IRE Transactions

ON RELIABILITY AND QUALITY CONTROL

Volume RQC-9
Follows PGRQC-16

APRIL,1960

TABLE OF CONTENTS

PUBLISHED BY THE

PROFESSIONAL GROUP ON RELIABILITY AND QUALITY CONTROL

IRE PROFESSIONAL GROUP ON RELIABILITY AND QUALITY CONTROL

The Professional Group on Reliability and Quality Control is an organization, within the framework of the IRE, of members with principal professional interest in Reliability and Quality Control. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of prescribed fees.

Annual Fee: $2.00

IRE TRANSACTIONS®
on Reliability and Quality Control

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Quality Control, 1 East 79th Street, New York 21, New York. Responsibility for the contents rests upon the authors, and not upon the IRE, the Group or its members. Individual copies available for sale to IRE-PGRQC members at $2.35, to IRE members at $3.50 and to nonmembers at $7.05.

PRINTED IN U.S.A.

CALCULATION OF AVERAGE FAILURE-TO-FAILURE TIME OF EQUIPMENT*

M. A. SINITSA†

INTRODUCTION

The average failure-to-failure time is the most convenient criterion of the reliability of various radio and electronic equipment. It is known that this value varies considerably throughout operation, especially in the case of airborne equipment. At the present time, calculation of reliability is widely made on the basis of the exponential law, the use of which is determined by the constancy of the average failure-to-failure time, or, and this is exactly the same thing, the constancy of the average frequency of failures during operation. It is of interest to find the law of change of the average failure-to-failure time throughout the whole period of operation of equipment, expressed through the laws of probability distribution of the failure-free operation time of components.

LAWS OF PROBABILITY DISTRIBUTION OF THE OPERATION TIME OF EQUIPMENT UNTIL THE FIRST FAILURE

The probability of damage to equipment not provided with reserve devices may be expressed through the probability of damage of components, as shown in Siforov's paper [1] to be

$$G = \lambda - \prod_{i=1}^{n} (1 - F_i)$$

where

F_i = the probability of damage to the ith component,

G = the probability of damage to the equipment,

n = the total number of components in the equipment.

It is here considered that the values of F_i are statistically independent. Considering the values of F_i as functions of the probability distribution of the failure-free operation time of the components, it is possible to determine the function of the probability distribution of occurrence of the first failure of the equipment as

$$G(t) = 1 - \prod_{i=1}^{n} [1 - F_i(t)]. \tag{1}$$

If all the components of the equipment possess the same laws of probability distribution, then

$$G(t) = 1 - [1 - F(t)]^n. \tag{2}$$

The probability density function of the time of occurrence of the first failure of the equipment can be determined by differentiating (1) in t

$$g(t) = \sum_{i=1}^{n} f_i(t) - \sum_{i=1}^{n} \left\{ f_i(t) \sum_{\substack{j=1 \\ j \neq i}}^{n} F_j(t) \right\} + \sum_{i=1}^{n} \left\{ f_i(t) \sum_{\substack{j,k=1 \\ j \neq k \neq i}}^{n} F_j(t) F_k(t) \right\} - \ldots \tag{3}$$

$$+ (-1)^{n-1} \sum_{i=1}^{n} \left\{ f_i(t) \prod_{\substack{j=1 \\ j \neq i}}^{n} F_i(t) \right\}$$

where $g(t) \neq \frac{dG(t)}{dt}$, $f(t) = \frac{dF(t)}{dt}$.

These expressions make it possible in principle to find the laws of probability distribution of the operation time of the equipment until the first failure on the basis of the known laws of probability distribution of failure-free operation of the components. However, it is extremely difficult to use such formulas in practice, because even with $n = 4$, (3) contains 32 terms. True, in the case when $F_i(t) \ll I$ it is possible in (1) and (3) to restrict ourselves to only the first sums containing the terms $F(t)$ and $f(t)$ in the first degree, which is what was done in the paper of Siforov [1]. However, such simplification is not always acceptable as it is actually necessary here to fulfill the condition $nF(t) \ll 1$ and not merely $F(t) \ll 1$. With a large number of components and insufficiently low values of $F(t)$ the former condition may be violated. The problem arises of approximating (1)

*This Paper was Presented at the Fifth Natl. Symp. on Reliability and Quality Control, Philadelphia, Pa., January 12-14, 1959.

†USSR.

and (3) by functions more convenient for computation and ensuring sufficient accuracy in engineering calculations. With this aim in view it is expedient to allow certain simplifications in posing the problem. First of all, all the n components are grouped into a small number (m) of subgroups, presuming that the components of each separate subgroup possess the same laws of distribution. Then (1) may be presented as

$$G(t) = 1 - [1 - F_1(t)]^{n_1}[1 - F_2(t)]^{n_2}\ldots[1 - F_m(t)]^{n_m}. \tag{4}$$

Most of the widely used components of radio equipment have an F(t) value above 1000 hours, not exceeding 0.01 to 0.02 (3) (*i.e.*, for values of t above 1000 hours, F(t) is not more than 0.01 or 0.02). Only in the case of such components as magnetrons, klystrons and certain types of UHF electron-vacuum devices do the functions F(t) noticeably exceed the indicated values. However, as the number of such components in equipment is usually small, it is possible to replace all the binominals in (4) by an approximated expression of the type $e^{-nF(t)}$. Then (4) may be given as

$$G(t) = 1 - \exp\left\{- \sum_{i=1}^{m} n_i F_i(t)\right\}. \tag{5}$$

Differentiating in t, we obtain the probability density function of the time of occurrence of the first failure of the equipment:

$$g(t) = \sum_{i=1}^{m} n_i f_i(t) \exp\left\{- \sum_{i=1}^{m} n_i F_i(t)\right\}. \tag{6}$$

The value of the relative error due to the replacement of a binominal of the type $(1 - F)^n$ by the exponential e^{-nF} can be evaluated from the following expression:

$$\delta = \frac{e^{-nF} - (1 - F)^n}{(1 - F)^n}. \tag{7}$$

From this formula the curves of the relative error for various numbers of components in group n are plotted and are shown in Fig. 1. The graphs pertain only to (5) and cannot be unconditionally applied to (6), as differentiation in itself of the approximated function can produce considerable errors.

We shall now evaluate the magnitude of the relative error which arises when (6) is applied to one group of components. For this we first determine the exact value of the function g(t)

$$g(t) = nf(t)[1 - F(t)]^{n-1}$$

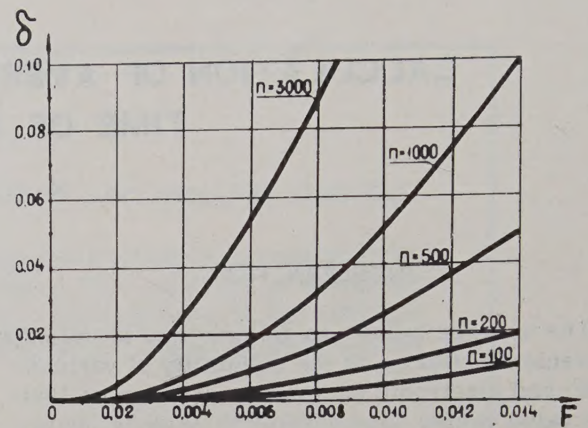


Fig. 1—Relative error in exponential approximation to power of a binomial.

and the approximate value,

$$g^*(t) = nf(t)e^{-nF(t)},$$

and then by formula

$$\gamma = \frac{g^*(t) - g(t)}{g(t)}$$

we find

$$\gamma = \frac{e^{-nF(t)}}{[1 - F(t)]^{n-1}} - 1. \tag{8}$$

Reducing (7) to a similar form

$$\delta = \frac{e^{-nF(t)}}{[1 - F(t)]^n} - 1$$

it can be noted that the relative errors of (5) and (6) will be of the same order. The graphs in Fig. 2 which are plotted from (8) clearly confirm this.

Thus, our approximation of binominals by exponential expressions is quite acceptable as regards the accuracy of the result obtained and sufficiently convenient for practical calculations.

Eqs. (5) and (6) could also have been derived directly from Poisson's Law

$$G(k) = \frac{\beta^k}{k!} e^{-\beta},$$

where G(k) = the probability of damage to K components of a total number of n. The parameter β equals nF, where F is the probability of damage to a component. As applied to the conditions of our problem, Poisson's Law can be presented in the following form:

$$G(k) = \frac{(nF)^k}{k!} e^{-nF}.$$

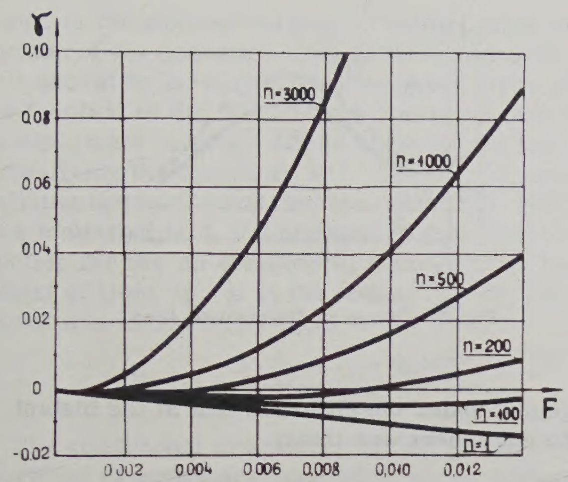


Fig. 2—Relative error in probability formula resulting from use of exponential approximation.

Usually Poisson's formula is used in cases where F = const., *i.e.*, under constant experimental conditions. This means that the probability of damage of K components out of n is a function of only n and K. However, we are justified in considering several series of experiments. The value of F, while remaining constant within one series of experiments, can change according to a certain law on crossing from one series to another. Each series of experiments can be made to correspond to a certain value of time t, since $F = F(t)$. This means that the magnitude G is a function not only of n and K but also of time t: $G = G(n,k,t)$. Supposing K and n to be constant for all series of experiments, we obtain the value of G in the form of a function of time alone.

$$G(t) = \frac{[nF(t)]^k}{k!} e^{-nF(t)} .$$

In the given case we are concerned with the probability of failure of the system in the time t, *i.e.*, the probability that in the time t at least one of the components will be damaged.

$$P(t) = \sum_{k=1}^{n} G_k(t) .$$

This value can be calculated more simply through the probability of the opposite occurrence

$$P(t) = 1 - Q(t)$$

$$Q(t) = G_o(t) = e^{-nF(t)},$$

where

$$Q(t) = G_o(t) = e^{-nF(t)} ,$$

i.e., we have obtained an expression for the probability of damage in a certain group of components possessing the distribution function F(t). The reliability of the equipment is determined by the product of these values,

$$Q_a(t) = \exp\left\{- \sum_{i=1}^{m} n_i F_i(t)\right\},$$

from which may be obtained the probability distribution function of the operation time of the equipment until the first failure determined by (5).

$$G_a(t) = 1 - \exp\left\{- \sum_{i=1}^{m} n_i F_i(t)\right\}.$$

The use of Poisson's asymptotic formula in the given case is subject to the same conditions which were accepted above, namely, a sufficiently low value of F(t) and high value of n. This is why evaluation of the error incurred when using Poisson's formula can also be carried out by the method suggested by B. V. Gnedenko [2].

On the basis of the obtained laws of probability distribution it is possible to find the average time of operation until the first failure or, and this amounts to the same thing, the mathematical expectancy of the failure-free operation time.

$$\overline{t}_1 = \int_0^{\infty} t g(t) dt .$$

Replacing g(t) in this formula by (6), we obtain

$$\overline{t}_1 = \sum_{i=1}^{m} n_i \int_0^{\infty} t f_i(t) \exp\left\{- \sum_{i=1}^{m} n_i F_i(t)\right\} dt . \qquad (9)$$

Dispersion of the time of operation until the first failure can be determined by

$$D[t_1] = \sum_{i=1}^{m} n_i \int_0^{\infty} (t - \overline{t}_1)^2 f_i(t) \exp\left\{- \sum_{i=1}^{m} n_i F_i(t)\right\} dt. \quad (10)$$

Thus, having the laws of probability distribution of damage of components of the equipment f(t) or F(t), it is possible to find the laws of the probability distribution of the operation time of the equipment until the first failure [g(t) and G(t)] and the basic numerical characteristics of these laws, the mathematical expectancy of the dispersion.

THE LAW OF CHANGE OF THE AVERAGE FAILURE-TO-FAILURE TIME

The average failure-to-failure time of equipment can be determined in several ways. Strict theoretical determination of this value demands an evaluation of the probability of replacement of one of the components of a group after t_1 hours of operation of the equipment. Next it is necessary to evaluate the conditional laws of the probability distribution of the time of damage of the equipment under the condition that one or another of the components is replaced. On the basis of these laws, there can be determined a series of conditional values of the average failure-to-failure time $\overline{t}_2$. Finally, on the basis of the conditional values of $\overline{t}_2$, with allowance for the specific weight of each group of components in the equipment, it is possible to find the mathematical expectancy of the operation time between the first and second failures. This method of finding $\overline{t}_2$ is extremely laborious. Taking into consideration that the components in modern radio and electronic equipment are numbered in thousands and sometimes in tens and hundreds of thousands, it is possible to overlook the effect of the replacement of one or several components on the general law of probability distribution. Then, the method of approximate calculation of the average time of operation between the first and second failures will boil down to the instant $\overline{t}_1$ is taken as the point of departure from which we determine the mathematical expectancy of the operation time until the second failure $\overline{t}_2$. When finding the laws of distribution for equipment in this case, we do not use the function f(t), but the normalized probability densities calculated under the supposition that at the instant $\overline{t}_1$ all the components were intact.

$$\phi(\theta) = \frac{f(\overline{t}_1 + \theta)}{1 - \int_0^{\overline{t}_1} f(t)dt}. \tag{11}$$

In Fig. 3 this conditional law of distribution is shown by the broken line.

On the basis of the function f(t) known for each group of components, it is possible to find the conditional probability densities using (6).

$$w(\theta) = \sum_{i=1}^{m} n_i\phi_i(\theta)\exp\left\{- \sum_{i=1}^{m} n_i\int_0^{\theta} \phi_i(\theta)d\theta\right\}. \tag{12}$$

Next, it is not hard to find the mathematical expectance of the failure-free operation time of the

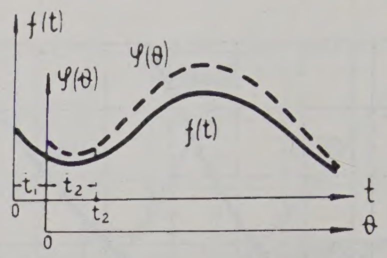


Fig. 3—Change in distribution laws.

equipment under the condition that at the instant $\overline{t}_1$ the equipment was intact.

$$\overline{t}_2 = \sum_{i=1}^{m} n_i \int_0^{\infty} \theta\phi_i(\theta)\exp\left\{- \sum_{i=1}^{m} n_i \int_0^{\theta} \phi_i(\theta)d\theta\right\} d\theta. \tag{13}$$

When determining the average time of operation between the second and the third failure, $\overline{t}_3$, it is similarly possible to normalize the function f(t) with respect to the instant of time $\overline{t}_1 + \overline{t}_2$

$$\psi(\xi) = \frac{f(\xi + \overline{t}_1 + \overline{t}_2)}{1 - \int^{\overline{t}_1+\overline{t}_2} f(t)dt}$$

and to find the new conditional law of probability distribution, $v(\xi)$.

Having obtained in this way a series of values for $\overline{t}_k$, it is possible to plot a graph of the law of change of the average failure-to-failure time depending on the time of operation of the equipment.

The instant of time $\overline{t}_k$ need not necessarily be associated with the instant of occurrence of the kth failure. It can obviously be selected at random on the time base. Then the distribution density $w(\xi,t_k)$ must be regarded as the function of the two arguments, ξ and t_k, where t_k is the instant with respect to which the laws of probability distribution of the component failure f(t) are normalized, while ξ is the current value of time from the point of departure t_k. The mathematical expectancy of the operation time between two consecutive failures can, in this case, be expressed by

$$\overline{t}_{k+1} = \frac{k}{Q(t_k)} \int_0^{\infty} \xi\, w(t_k + \xi)d\xi. \tag{14}$$

Here

$$Q(t_k) = 1 - \int_0^{t_k} g(t)dt. \tag{15}$$

Thus, it seems possible to obtain not simply a series of discrete values of $\overline{t}_k$, but a law of

change of the average failure-to-failure time as a function of the operation time of the equipment.

It should be mentioned that the probability density function of the failure-free operation time of the equipment $w(\xi,t_k)$ can be obtained not only by normalizing the functions $f(t)$, but also by normalizing the initial distribution density of $g(t)$. As a matter of fact, the probability distribution function for the ith component, counted from any instant of time t_k, if at the instant t_k the component was in order, is equal to

$$F_{i_k}(\xi) = \frac{F_i(t_k + \xi) - F_i(t_k)}{1 - F_i(t_k)} .$$

The conditional probability of failure-free operation of equipment consisting of m groups of components can be expressed, using the function $F_k(\xi)$, as

$$Q_k(\xi\ t_k) = \prod_{i=1}^{m} [1 - F_{i_k}]^{n_i} = \prod_{i=1}^{m} \left[\frac{1 - F_i(t_k+\xi)}{1 - F_i(t_k)}\right]^{n_i}$$

$$= \frac{Q(t_k + \xi)}{Q(t_k)}$$

where $Q(t)$ gives the reliability of the equipment.

Taking into account that

$$w(\xi,t_k) = -\frac{d\theta_k(\xi)}{d\xi} ,$$

we obtain

$$w(\xi,t_k) = \frac{g(t_k + \xi)}{Q(t_k)} .$$

It can be shown that this conclusion is also correct when the exponential approximation is employed.

$$Q_k(\xi) = \exp\left\{-\sum_{i=1}^{m} n_i[F_i(t_k + \xi) - F_i(t_k)]\right\}$$

$$= \exp\left\{-\sum_{i=1}^{m} n_i \int_{t_k}^{t_k+\xi} f_i(t)dt\right\}.$$

The conditional probability density of failure of the equipment is now calculated on the assumption that at the instant t_k, the equipment was in order.

$$w(\xi,t_k) = -\frac{dQ(\xi)}{d\xi} = \sum_{i=1}^{m} n_i f_i(t_k + \xi)\exp$$

$$\left\{-\sum_{i=1}^{m} n_i \int_{0}^{t_k+\xi} f_i(t)dt + \sum_{i=1}^{m} n_i \int_{0}^{t_k} f(t)dt\right\} . \qquad (16)$$

As $t_k + \xi = t$, then

$$w(\xi,t_k) = \frac{\sum_{i=1}^{m} n_i f_i(t)\exp\left\{-\sum_{i=1}^{m}\int_{0}^{t} f_i(t)dt\right\}}{\exp\left\{-\sum_{i=1}^{m} n_i \int_{0}^{t_k} f(t)dt\right\}} , \qquad (17)$$

or

$$w(t) = \frac{g(t)}{1 - \int_{0}^{t_k} g(t)dt} \quad \text{when} \quad t_k \leqslant t < \infty, \qquad (18)$$

which needs to be shown.

Therefore, in order to determine the law of change of the average failure-to-failure time it is first necessary to find the conditional distribution densities for the components or for the equipment. We are of the opinion that normalization of the laws of distribution for equipment is the more convenient operation.

CONCLUSION

The report produced a number of asymptotic expressions for the laws of probability distribution of the operation time of equipment until the first failure. An evaluation was made of the errors incurred when crossing over from exact to approximate expressions and it was shown that, when applied to modern equipment, the errors of approximated formulas did not exceed several per cent.

The report dealt with the methods of approximate calculation of the law of change of the average failure-to-failure operation time of equipment, from any optionally chosen instant of time t_k. It was shown that normalization of the laws of distribution for components can be replaced by the equivalent, but less laborious, operation of normalizing the initially obtained probability density function of failure of equipment.

REFERENCES

[1] V. I. Siforov, "On the methods of calculating the reliability of operation of systems containing a large number of components," *Izv. Akad. Sci. USSR*, no. 6; 1954.

[2] B. V. Gnedenko, "Course of the Theory of Probability," GITTL, m. 1954.

[3] I. I. Morozov, "Reliability of operation of radioelectric equipment," *Radioelectronnaya Promishlyennost*, no. 3; 1958.

ON RESERVATION BY REPLACEMENT METHOD*

M. A. SINITSA†

Summary —Consideration of substitution techniques for reliable design is usually based upon statistical independence of failures of operating and spare components.

A method of design of a reliable system, consisting of many operating and spare components, for the case of statistical relation between failures of operating and spare components is considered. In a number of cases more reliable operation of systems can be achieved by providing substituting reservation, instead of permanent switching, to a spare (or "hot" reserve) component. Effective usage of a single spare component substituted for several operating components is considered.

The reservation of radioelectronic equipment by a replacement method, due to a number of operation and economic advantages, finds ever increasing application in engineering. The reliability of systems with such a reservation cannot be evaluated by the methods described by Moskowitz,[1] Levin,[2] and Welber, *et al.*,[3] as they are all based on the assumption of statistical independence of failure of basic and reserve components. When reserving by the replacement method, each reserve component is put into service after failure of the basic one. This means that the probability of failure of the reserve equipment depends upon the probability of failure of the basic component. In this report an attempt is made to consider a probability evaluation method of complex systems, consisting of many operating and reserve components, on condition that failure of the reserve components is statistically dependent on failure of the basic components, or reserve ones which have been previously placed in service.

Let us assume that the system consists of K identical operating components, and m similar reserve ones. Each of the reserve components can replace any failed operating component or reserve one which has been previously placed in service. The effect of switching devices on reliability will not be taken into account here, lest the problem should be complicated at the first stage. To determine probability of failure of such a system, we shall use the method described elsewhere by the author.[4]

At first, let us consider the solution of a simple problem: when there are K operating components, and one reserve one which can replace any of the operating components. In this case failure of the system (during time T) may be the result of either of two mutually exclusive events:

1) Failure of not less than one operating component and a reserve one.

2) Failure of not less than two operating components with the reserve components operating properly.

Let us divide the greatest possible operating time of components without failure into W intervals, of duration $\Delta\tau$. Let us take as hypotheses:

a) For the first event, H_1: failure of the system during interval i.

b) For the second event, H_2: failure of not less than two operating components during interval j.

The probability of system failure during time T can be determined with the aid of the formula of the total probability:

$$P_k^{(1)}(T) = \sum_{i=1}^{W} \Delta P(H_1^i)p(T/\tau_i) + \sum_{j=1}^{W} \Delta P(H_2^j)q(T/\tau_j),^5 \tag{1}$$

where

$\Delta P(H_1^i)$ = probability of failure of not less than one component during interval i [probability of hypothesis a)]

*This Paper was Presented at the Fifth Natl. Symp. on Reliability and Quality Control, Philadelphia, Pa., January 12-14, 1959.

†USSR.

[1] F. Moskowitz, and I. B. McLean, "Some reliability aspects of systems design," IRE TRANS. ON RELIABILITY AND QUALITY CONTROL, no. PGRQC-8, pp. 7-35; September, 1956.

[2] B. R. Levin, "Increase of system reliability by reservation," *Elektrosviaz,* no. 11; 1956.

[3] I. Welber, H. W. Evans, and G. A. Pullis, *Bell Sys. Tech. J.* vol. 34, p. 473-510; 1955.

[4] M. A. Sinitsa, "Reservation methods of radioequipment," *Elektrosviaz,* no. 7; 1958.

[5] Note that the notation $p(T\backslash\tau_i)$ apparently does not mean p(x) where x is the quotient $T\backslash\tau_i$; but is rather a way of indicating functional dependence on both T and τ_i, as though written $p(T\backslash\tau_i$, or $p(T;\tau_i)$, or $p(T,\tau_i)$. (Ed.)

$\Delta P(H_2^j)$ = probability of failure of not less than two components during interval j [probability of hypothesis b)]

$p(T/\tau_i)$ = conditional probability of failure of the reserve component during time T. It is obtained on the supposition that the first component of the operating system has failed during interval i

$q(T/\tau_j)$ = conditional probability of the reserve component successful operation during time T. It is obtained on the supposition that the reserve component has been placed in service during interval j.

Increasing the number of intervals to infinity ($w \to \infty$) and passing over to limits, we obtain

$$P_k^{(1)}(T) \quad \int_0^T dP(H_1)p(T/\tau) + \int_0^T dP(H_2)q(T/\tau). \qquad (2)$$

Let us consider the expressions within the integral. The differential of the probability of hypothesis a) is the probability that failure will occur during interval dt at τ:

$$dP(H_1) = r_k^{(1)}(t)dt, \qquad (3)$$

where $r_k^{(1)}(t)$ is the density of the probability of failure in the system of not less than one operating component from the total number of K components.

As is shown elsewhere by the author,[6] $r_k^{(1)}(t)$ can be expressed in terms of density of the probability of failure of components f(t), according to the formula

$$r_k^{(1)}(t) = kf(t)[\int_t^\infty f(t)dt]^{k-1}.$$

Similarly

$$dP(H_2) = r_k^{(2)}(t)dt, \qquad (4)$$

where $r_k^{(2)}(t)$ is the density of the probability of failure of the system, assuming that not less than two components have failed. $r_k^{(2)}(t)$ can also be expressed by the functions f(t):[6]

$$r_k^{(2)}(t) = k(k-1)f(t)[\int_t^\infty f(t)d\tau]^{k-2}\int_0^t f(\tau)d\tau.$$

The conditional probability of failure of the reserve component, and the conditional probability of its successful operation, are determined by the equations

$$\left.\begin{aligned} p(T/\tau) &= \int_0^T f(\tau,t)dt \\ q(T/\tau) &= \int_T^\infty f(\tau,t)dt \end{aligned}\right\} \qquad (5)$$

In these equations the function within the integral will have various expressions at different time intervals

$$f(\tau,t) = \begin{cases} f'(t) & \text{when } t < \tau, \\ f''(t) & \text{when } t \geqslant \tau. \end{cases} \qquad (6)$$

Taking into account (3), (4) and (5), the expression (2) will be:

$$P_k^{(1)}(T) = \int_0^T r_k^{(1)}(\tau)\left\{\int_0^T f(\tau,t)dt\right\}d\tau + \int_0^T r_k^{(2)}(\tau)[\int_T^\infty (\tau,t)dt]d\tau. \qquad (7)$$

The formula (7) allows us to evaluate the probability of failure of the system consisting of K operating components and 1 reserve component, which have been placed in service after failure of the operating component. It is a most common formula for a system with 1 reserve component in which failure can be statistically both a dependent event and an independent one. For example, the expression for the probability of failure of a system with a hot reserve component (statistically independent component) can be obtained as a particular case of (7). Indeed, as $f(\tau, t) = f(t)$, the expressions within the integral for both cases of (7) decompose into two functions of independent variables

$$P_k^{(1)}(T) = \int_0^T r_k^{(1)}(\tau)d\tau \int_0^T f(t)dt + \int_0^T r_k^{(2)}(\tau)d\tau \int_T^\infty f(t)dt.$$

Considering that

$$\int_0^T r_k^{(j)}(\tau)d\tau = \sum_{i=j}^k C_k^i[\int_0^T f(\tau)d\tau]^i[\int_T^\infty f(\tau)d\tau]^{k-i};$$

$$\int_0^T f(\tau)d\tau = p \quad \text{and} \quad \int_T^\infty f(\tau)d\tau = q, \qquad (8)$$

the last expression will be

$$P_k^{(1)}(T) = (C_k^1 pq^{k-1} + C_k^2 p^2 q^{k-2} + \ldots + p^k)p + (C_k^2 p^2 q^{k-2} + C_k^3 p^3 q^{k-3} + \ldots + p^k)q.$$

[6]M. A. Sinitsa, "Reservation of radioelectronic equipment," *Radioelektronnaya Promushlennoct*, no. 5; 1958.

After rather simple transformation we obtain:

$$P_k^{(1)}(T) = (C_k^1 + C_k^2)p^2q^{k-1} + (C_k^2 + C_k^3)p^3q^{k-2} + \ldots + (C_k^{k-1} + C_k^k)p^kq + p^{k+1}.$$

Using the property of binomial coefficients we can write:

$$C_k^1 + C_k^2 = C_{k+1}^2, \quad C_k^2 + C_k^3 = C_{k+1}^3, \ldots, \quad C_k^{k-1} + C_k^k = C_{k+1}^k,$$

then

$$P_k^{(1)}(T) = C_{k+1}^2 p^2 q^{k-1} + C_{k+1}^3 p^3 q^{k-2} + \ldots + C_{k+1}^1 p^k q + p^{k+1},$$

or

$$P_k^{(1)}(T) = \sum_{i=2}^{k+1} C_{k+1}^i p^i q^{k+1-i}. \tag{9}$$

Thus, we have obtained the expression which completely agrees with the formula of the probability of system failure with one hot reserve component, previously described.[6]

Let us consider the solution of the same problem for the case of two reserve components in a system with all other conditions being retained.

For convenience a system without reserve components will be referred to as a "basic system." Failure of the system during time T is a case of the sum of three noncoincident events:

1) Failure in the basic system of not less than 1 operating component and 2 reserve ones which have been placed in service in succession.

2) Failure in the basic system of not less than 2 operating components and a reserve one, with the other reserve component operating properly.

3) Failure in the basic system of not less than 3 operating components with 2 reserve ones operating properly.

The hypotheses of these three items are:

a) the hypothesis of the first item H_1 is failure of the basic system and the first reserve component within a time interval;

b) the hypothesis of the second item H_2 is failure in the basic system of not less than 2 components and a reserve one within an elementary time interval;

c) the hypothesis of the third item is failure in the basic system of not less than three components within the one of an elementary time interval with the reserve component operating properly.

The probability of these hypotheses are:

$$\left.\begin{aligned} dP(H_1) &= r_k^{(1)}(\tau)\left[\int_0^T f(\tau,t)dt\right]d\tau, \\ dP(H_2) &= r_k^{(2)}(\tau)\left[\int_0^T f(\tau,t)dt\right]d\tau, \\ dP(H_3) &= r_k^{(3)}(\tau)\left[\int_T^\infty f(\tau,t)dt\right]d\tau \end{aligned}\right\} . \tag{10}$$

The probability densities, $r_k^{(2)}(t)$ and $r_k^{(3)}(t)$, can be evaluated, as is shown by the author,[6] according to the functions $f(t)$

$$r_k^{(i)}(t) = \frac{k!}{(k-i)!(i-1)!} f(t)\left[\int_0^t f(\tau)d\tau\right]^{i-1}\left[\int_t^\infty f(\tau)d\tau\right]^{k-i}. \tag{10a}$$

In the same way, the probability of system failure with two reserve components can be determined, according to the formula of the total probability

$$P_k^{(2)}(T) = \int_0^T dP(H_1)p(T/\xi) + 2\int_0^T dP(H_2)q(T/\lambda) + \int_0^T dP(H_3)q(T/\theta), \tag{11}$$

where

$p(T/\xi)$ = conditional probability of failure of the second reserve component. It is obtained on the supposition that the second reserve component has been placed in service at the moment $\xi \geqslant \tau$, that is, when the second operating component or the first reserve one has failed.

$q(T/\lambda)$ = conditional probability of the reserve component proper operations. It is obtained on the supposition that the reserve component has been placed in service at the moment $\lambda \geqslant \tau$.

$q(T/\theta)$ = conditional probability of the reserve component proper operation. It is obtained on the supposition that the reserve component has been placed in service at the moment θ.

Taking into account that

$$p(T/\xi) = \int_0^T f(\xi,t)dt,$$

$$q(T/\lambda) = \int_T^\infty f(\lambda,t)dt,$$

$$q(T/\theta) = \int_T^\infty f(\theta,t)dt,$$

and making allowance for (10), formula (11) will be:

$$P_k^{(2)}(T) = \int_0^T r_k^{(1)}(\tau)d\tau \int_0^T f(\tau,\xi)d\xi \int_0^T f(\xi,t)dt$$

$$+ 2\int_0^T r_k^{(2)}(\tau)\left\{\int_0^T f(\tau,t)[\int_T^\infty f(\lambda,t)dt]dt\right\} d\tau$$

$$+ \int_0^T r_k^{(3)}(\tau)[\int_T^\infty f(\tau,t)dt][\int_T^\infty f(\theta,t)dt]d\tau. \quad (12)$$

Here again, the evaluation of conditional probabilities is performed for two time intervals. In this case density probabilities will be

$$f(\xi,t) = \begin{cases} f'(t) & \text{when } t < \xi, \\ f''(t) & \text{when } t \geqslant \xi, \end{cases}$$

$$f(\lambda,t) = \begin{cases} f'(t) & \text{when } t < \lambda, \\ f''(t) & \text{when } t \geqslant \lambda. \end{cases}$$

Eq. (12) is given in a most common form. To make it applicable in practice we should specify the succession of integration. The first item of the right-hand side of the equation contains the probabilities of two groups of events, for which the succession of integration differs one from the other. The first group of events is the failure of exactly one component in the operating system and each of the reserve ones which has been placed in service in succession. The second group of events is the failure of not less than two operating components and reserve ones which have replaced each failed component. In view of the aforesaid, (12) will be:

$$P_k^{(2)}(T) = \int_0^T h_k^{(1)}(\tau)\left\{\int_0^T f(\tau,\xi)[\int_0^T f(\xi,\lambda)d\lambda]d\xi\right\}d\tau$$

$$+ \int_0^T r_k^{(2)}(\tau)[\int_0^T f(\tau,\xi)d\xi]^2 d\tau$$

$$+ 2\int_0^T r_k^{(2)}(\tau)\left\{\int_0^T f(\tau,t)dt[\int_T^\infty f(\tau,t)dt]\right\} d\tau$$

$$+ \int_0^T r_k^{(3)}(\tau)[\int_T^\infty f(\tau,t)dt]^2 d\tau. \quad (13)$$

Here $hk^{(1)}(\tau)$ = density probabilities of failure of exactly one component out of K operating ones, and $hk^{(1)}(\tau)$ is determined according to the formula obtained elsewhere by the author.[6]

$$h_k^{(i)}(t) = C_k^i f(t)[\int_0^t f(\tau)d\tau]^{i-1}[i - k\int_0^t f(\tau)d\tau]$$

$$[\int_T^\infty f(\tau)d\tau]^{k-i-1} \quad (14)$$

Now let us follow the change of (13) as the reserve environment is changed. If the reserve components do not wear out before being placed in service, then

$$f(\tau,\xi) = \begin{cases} 0 & \text{when } \xi < \tau, \\ f(\xi) & \text{when } \xi \geqslant \tau, \end{cases} \qquad f(\xi,\lambda) = \begin{cases} 0 & \text{when } \lambda < \xi, \\ f(\lambda) & \text{when } \lambda \geqslant \xi, \end{cases}$$

and (13) will be

$$P_k^{(2)}(T) = \int_0^T h_k^{(1)}(\tau)\left\{\int_0^{T-\tau} f(\xi)[\int_0^{T-\tau-\xi} f(\lambda)d\lambda]\, d\xi\right\}d\tau$$

$$+ \int_0^T r_k^{(2)}(\tau)[\int_0^{T-\tau} f(\xi)d\xi]^2 d\tau + 2\int_0^T r_k^{(2)}(\tau)\left\{\int_0^{T-\tau} f(t)dt\right.$$

$$\left.[\int_{T-\tau}^\infty f(t)dt]\right\}d\tau + \int_0^T r_k^{(3)}(\tau)[\int_{T-\tau}^\infty f(t)dt]^2 d\tau. \quad (15)$$

If the reserve components are hot, that is, $f(\xi,t) = f(t)$, $f(\lambda,\xi) = f(t)$, . . . , then the expressions within the integral in (13) decompose into function products of independent variables:

$$P_k^{(2)}(T) = \int_0^T h_k^{(1)}(\tau)d\tau[\int_0^T f(\xi)d\xi][\int_0^T f(\lambda)d\lambda] + \int_0^T r_k^{(2)}(\tau)d\tau$$

$$[\int_0^T f(\xi)d\xi]^2 + 2\int_0^T r_k^{(2)}(\tau)d\tau \int_0^T f(t)dt[\int_T^\infty f(t)dt]$$

$$+ \int_0^T r_k^{(3)}(\tau)d\tau[\int_T^\infty f(t)dt]. \quad (16)$$

In this case

$$\int_0^T h_k^{(1)}(\tau)d\tau = kpq^{k-1},$$

$$\int_0^T r_k^{(2)}(\tau)d\tau = \sum_{i=2}^{k} C_k^i p^i q^{k-i},$$

$$\int_0^T r_k^{(3)}(\tau)d\tau = \sum_{i=3}^{k} C_k^i p^i q^{k-i};$$

therefore the formula (16) will be:

$$P_k^{(2)}(T) = (C_k^1 pq^{k-1} + C_k^2 p^2 q^{k-2} + \ldots + p^k)p^2$$
$$+ 2(C_k^2 p^2 q^{k-2} + C_k^3 p^3 q^{k-3} + \ldots + p^k)pq$$
$$+ (C_k^4 p^4 q^{k-4} + \ldots + p^k)q^2.$$

After some transformations we obtain

$$P_k^{(2)}(T) = (C_k^1 + 2C_k^2 + C_k^3)p^3 q^{k-1} +$$
$$(C_k^2 + 2C_k^3 + C_k^4)p^4 q^{k-2} + \ldots$$
$$+ \left(C_k^{k-1} + 2C_k^k\right)p^{k+1}q + p^{k+2}.$$

In view of the property of binomial coefficients we obtain

$$C_k^1 + 2C_k^2 + C_k^3 = C_{k+2}^3,$$
$$C_k^2 + 2C_k^3 + C_k^4 = C_{k+2}^4,$$
$$\cdots\cdots\cdots\cdots$$
$$C_k^{k-1} + 2C_k^k = C_{k+2}^{k+1}.$$

Therefore

$$P_k^{(2)}(T) = C_{k+2}^3 p^3 q^{k-1} + C_{k+2}^4 p^4 q^{k-2} + \ldots$$
$$+ C_{k+2}^1 p^{k+1} q + p^{k+2}. \qquad (17)$$

And this is nothing but the probability of system failure consisting of K operating components and two reserve ones for case of hot reserve. Thus, (13) is common for evaluating the reliability of the system, consisting of K operating components and two reserve ones regardless of the environment of reserve components.

The above method can be applied for a system with three reserve components. The formula for determining the probability of system failure during time T should allow for the following groups of events:

1) Failure of exactly one component in the basic system and all the reserve ones which have been placed in service in succession instead of the failed component.

2) Failure of exactly two components in the basic system and all the reserve ones; one of the reserve components replacing the first failed component and the other two reserve components replacing the second failed component in succession.

3) Failure of not less than three components in the basic system and all the reserve ones, each of the reserve components replacing one of the operating components in succession.

4) Failure of not less than two components in the basic system and two reserve ones with one reserve component operating properly.

5) Failure of not less than three components in the basic system and a reserve one, the other two reserve components operating properly.

6) Failure of not less than four components in the basic system, all the reserve ones operating properly.

Such a plan of events allows us to determine the succession of integration in the formula of the probability of system failure having three reserve components.

$$P_k^{(3)}(T) = \int_0^T h_k^{(1)}(\tau)d\tau \int_0^T f(\tau,\xi)d\xi \int_0^T f(\xi,\lambda)d\lambda \int_0^T f(\lambda,\theta)d\theta$$
$$+ \int_0^T h_k^{(2)}(\tau)\left\{ \int_0^T f(\tau,\xi)\left[\int_0^T f(\xi,\lambda)d\lambda \right]d\xi \int_0^T f(\tau,\xi)d\xi \right\}d\tau$$
$$+ \int_0^T r_k^{(3)}(\tau)\left[\int_0^T f(\tau,t)dt\right]^3 d\tau + 3\int_0^T r_k^{(2)}(\tau)\left[\int_0^T f(\tau,t)dt\right]^2$$
$$\left[\int_T^\infty f(\tau,t)dt\right]d\tau + 3\int_0^T r_k^{(3)}(\tau)\left[\int_0^T f(\tau,t)dt\right]\left[\int_T^\infty f(\tau,t)dt\right]^2 d\tau$$
$$+ \int_0^T r_k^{(4)}(\tau)\left[\int_T^\infty f(\tau,t)dt\right]^3 d\tau. \qquad (18)$$

Here $h_k^{(1)}$, $hk^{(2)}$, $rk^{(2)}$, $rk^{(3)}$, $rk^{(4)}$ are determined according to the formula (14) and (10a).

If we apply the above method to the case of the hot reserve, (18) will be:

$$P_k^{(3)}(T) = C_{k+3}^4 p^4 q^{k-1} + C_{k+3}^5 p^5 q^{k-2} + \ldots$$
$$+ C_{k+3}^1 p^{k+2} q + p^{k+3} \qquad (19)$$

It is a particular case of the formula for determining the probability of system failure with a hot reserve.[6]

Using the same plan one can work out a formula for determining the probability of system failure system, consisting of K operating components and m reserve ones. This formula is very complicated, therefore there is no need to give it here. But we shall point out the basic groups of events that the formula should allow for:

1) Failure of not less than one of the total number of components in the basic system and all the reserve ones.

2) Failure of not less than two components out of K ones in the basic system and $m-1$ reserve ones with one reserve component operating properly.

3) Failure of not less than three components out of K ones and $m-2$ reserve ones with two reserve components operating properly.

.

.

.

$m + 1$) Failure of not less than $m + 1$ components out of K ones with all m reserve components operating properly.

Each of these groups, except the last, should be divided into subgroups, as was the case when $m = 2$ and $m = 3$. This subdivision will give the possibility to determine the succession of integration in the formula.

While solving the problem of determining the probability of system failure, some particular cases of the system containing K operating components and $1,2,3,\ldots, m$ reserve ones were considered. In these cases we did not set any limits on m. It is evident that the quantity of reserve components can be either more than the number of the operating components, or less than it.

Thus, the resulting formulas for evaluating the reliability of complex systems consisting of K operating components and m reserve ones, are the most common formulas, as they allow us to estimate the reliability of systems with a hot and cold reserve which can be in any environment. The main formulas[1,2,4] can be obtained as particular cases of these generalizing expressions.

At present in radio relay systems one or two reserve circuits per several operating ones are used. The same method of increasing reliability can be applied to a number of other radio-engineering systems. It would be of interest to consider the effectiveness of such a method of increasing reliability, and when it should be used. For this purpose let us consider the following relationship

$$W = \psi(q) ,$$

where q = reliability of operating and reserve components, and
W = reliability gain.

In general the reliability gain can be determined as the ratio of the probability of system failure without reserve, to the probability system failure with reserve

$$W = \frac{P(k)}{P(k,m)} .$$

It is assumed that the system consists of K operating components and m reserve ones. The probability of system failure without reserve, consisting of K components, can be determined by the equation

$$P(k) = 1 - q_1 q_2 \ldots q_k ,$$

and if all the operating components have equal reliability, the equation will be

$$P(k) = 1 - q^k .$$

The probability of system failure with reserve $P(k,m)$ is determined by various formulas depending upon the environment of the reserve component.

In case of cold reserve and $m = 1$, $P(K,m)$ can be evaluated by (7), and when $m = 2$, by (12). In cases of hot reserve the probability of failure is evaluated by a more common formula

$$P(k,m) = \sum_{i=1}^{k} C_{k+m}^{m+i}\, p^{m+i} q^{k-i} . \tag{20}$$

Here $p = 1 - q$.

Thus, when $q_1 = q_2 = \ldots = q_k$ and $m = 1$ when having hot reserves, we obtain:

$$W = \frac{1 - q^k}{\sum_{i=1}^{k} C_{k+m}^{m+i}\, p^{m+i} q^{k-i}} . \tag{21}$$

With the aid of this formula the curves $w = \psi(q)$, given in Figs. 1-3, are obtained. Analyzing these curves one can see that the use of one reserve component per several operating ones when $m = 1$ gives substantial gain in reliability only when $q > 0.7 - 0.8$ (see Fig. 1). This gain rises sharply as q increases. For example: when $k = 5$ and $q = 0.8$ then $W = 2$, and when $q = 0.95$ then $W \cong 7$. Certainly, when there are less operating components this gain in reliability will be still greater. Thus, when $k = 2$ and $q = 0.95$, then $W = 13.35$.

The term "multiple reservation" in this case means the ratio m/k. Here again, as in the case $K = 1$, the increase of multiple reservation

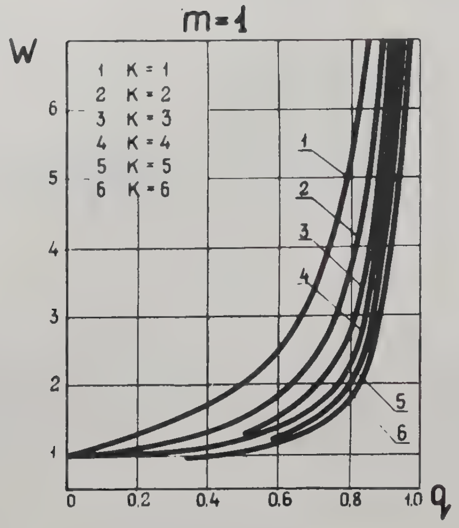


Fig. 1—Reliability gain. One "hot" reserve component.

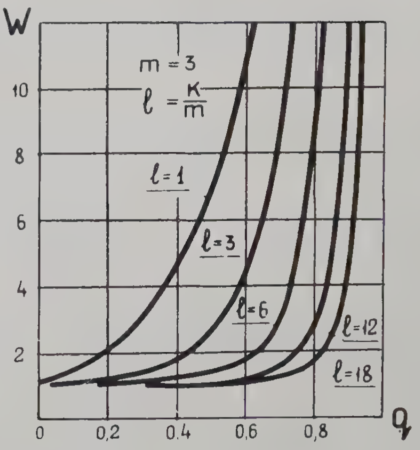


Fig. 3—Reliability gain. Three "hot" reserve components.

reduces the probability of system failure, that is, W increases. However, it should be pointed out that the gain in reliability rises substantially with the increase of M even when maintaining the multiple reservation $m/k = \text{const}$. For example, when $q = 0.8$ and $\frac{k}{m} = 6$ we find according to the diagrams (see Figs. 1, 2, and 3) that when $m = 1$ then $W = 1.75$; when $m = 2$ then $W = 3.7$; and when $m = 3$ then $W = 9$. For higher values of reliability of operating and reserve circuits $q > 0.8$ the increase of W is still more sharp. This is explained by the fact that each reserve component can replace any failed one out of the total K components, and even reserve ones which have been previously placed in service.

Now let us estimate what advantages we shall get when using cold reserve. For this purpose let us use the functions $P_{cold} = \phi_1(k)$, $P_{hot} = \phi_2(k)$, $W_{cold} = \psi_1(K)$, $W_{hot} = \psi_2(K)$ when $q = \text{const}$ and $m = 1$. The functions $P_{cold} = \phi_1(K)$, $P_{hot} = \phi_2(K)$ and $W_{hot} = \psi_2(K)$ are determined according to (7), (20) and (21) respectively and the function $W_{cold} = \psi_1(K)$ according to the formula

$$W_{cold} = \frac{P(k)}{P_{cold}(k,m)}\,, \tag{22}$$

where $P_{cold}(K,M)$ is also determined according to (7).

The curves which have been plotted after these formulas are given in Fig. 4. A curve of the

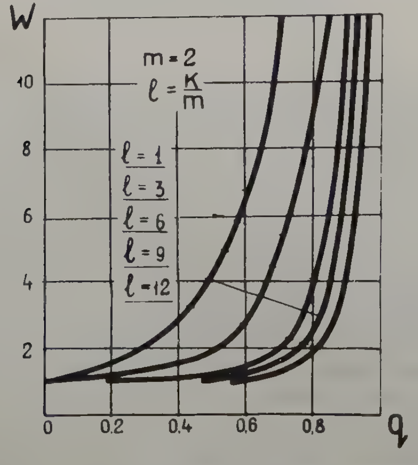


Fig. 2—Reliability gain. Two "hot" reserve components.

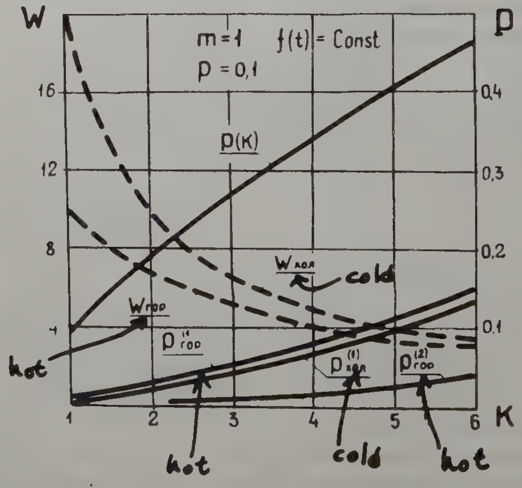


Fig. 4—Comparison of "hot" or "cold" reserve.

probability of system failure without reserve $P(k)$ as well as that of the probability of system failure when $m = 2$ for hot reserve $P^{(2)}_{hot}$ are given in Fig. 4 for comparison. The diagram is plotted for the case of the law of equal probability of component failure $f(t) = const$, when the probability of each component failure during time T is equal to $p = 0.1$. The diagram shows that cold reserve is much more preferable than hot reserve when there are one reserve component and few operating ones in the system. With the increase in the quantity of operating components in the system up to 5-6 this advantage decreases (see Table I).

TABLE I

K	2	3	4	5	6
$\frac{W_{cold}}{W_{hot}}$	1.47	1.28	1.21	1.17	1.10

It does not mean, however, that cold reserve should not be used at all. Cold reserve gives the possibility of economizing the energy of power supply and when two or more reserve components per 5-6 operating ones are used, a considerable gain in reliability is obtained.

One can solve a reciprocal problem with the aid of the diagram given in Fig. 4, that is, to determine the necessary quantity of reserve circuits for securing given reliability, if the quantity of operating circuits and their reliability are known. For example, it is necessary to secure the probability of system failure of not more than 0.05 in the system consisting of 6 operating circuits, each having the reliability of 0.9. Using the diagram (see Fig. 4) we determine that such reliability of the system can be secured when there are two reserve circuits.

Using the same principle it is possible to plot a diagram for various values of p and m by plotting curves for $P_{hot}(m)$ and $P_{cold}(m)$. These diagrams offer opportunity to solve the above problem for systems with various values of p and k. With the increase of m the significance of cold reserve also increases, and in some cases its use will allow us to decrease the number of reserve circuits while maintaining given reliability. Besides, it is considered to be expedient to use both hot and cold reserve when $m \geq 2$.

The problem of determining system reliability with cold reserve has been solved on supposition that the probabilities of failure of all the operating and reserve components are equal. It is the most common case in practice for the reserve components intended for carrying out the same functions as operating ones to consist of the same elements. In some cases, however, operating and reserve components can differ greatly from each other as to their reliability. For example, a component may or may not include electric vacuum devices. In this case, the method of determining reliability remains the same but formulas will be more complicated, as each item, before which a factor of the type C_k^m stands, must be represented as a sum of inequivalent components.

The method of evaluating reliability of complex systems consisting of several operating and reserve components given in this report offers opportunity to estimate the reliability of both available (present) systems and projected and developed ones, as the total reliability of a system is expressed by the laws of probabilities of its components failure, which can be obtained before. By changing the integration limits the system reliability during any operating period T can be evaluated. Moreover, using this method one can estimate the quantity of reserve components or circuits necessary to secure given reliability of the system.

The reliability estimation of a system with many operating and reserve components was given without accounting for the reliability effect of switching devices. The effect of switching devices on total system reliability can be easily taken into account for this case.[6] The accuracy of the evaluation of system reliability will be completely dependent upon the accuracy of initial data: the laws of probabilities of failure of components for given operating conditions.

SOME RESULTS OF MATHEMATICAL RELIABILITY THEORY*

B. R. Levin†

INTRODUCTION

The problem of reliability of electronic systems, (*i.e.*, their ability to fulfill specified functions under specified conditions) has two aspects: experimental and theoretical. The first concerns a statistical analysis of data obtained in the course of operational use, while the second concerns a mathematical analysis. The mathematical reliability theory is based on the statistical definition of component (or system) reliability, as probability of satisfactory operation during a given portion of time.

The mathematical reliability theory provides, for the engineer, the possibility of finding a rational method to predict reliability of the system by statistical analysis of experimental component reliability data.

Provision of these data is a necessary condition for practical application of the mathematical reliability theory.

Some results of the mathematical reliability theory obtained by the author are described in this paper.

RELIABILITY ANALYSIS FOR A FIXED FRACTION OF TIME

Consider a system as a matrix composed of $|A_{ij}|$ independent components. Each component is marked by letter A_{ij} where indexes i and j are numbers of row and column of the matrix, respectively. For a fixed interval of time T component A_{ij} is characterized by reliability P_{ij}. Thus, there is a reliability matrix $|P_{ij}|$ corresponding to the component matrix.

In each row the components are serially connected (whenever one of the operating components fails, the whole row breaks down). The rows are connected in parallel; thus the system breaks down only in case of failure of all the rows.

*This paper was presented at the Fifth Natl. Symp. on Reliability and Quality Control, Philadelphia, Pa., January 12-14, 1959.

†USSR.

Redundancy in the system with the corresponding matrix of $m \times n$ rows and columns can be provided by splitting the matrix into $r \leqslant n$ submatrices of $m_k \times s_k$ (for $k = 1, 2, \ldots r$) rows and columns serially connected. This means that $\sum_{k=1}^{r} S_k = n$.

Reliability of the system will be given by

$$P = \prod_{k=1}^{r} \left\{1 - \prod_{i=1}^{m_k} \left(1 - \prod_{j=S_{k-1}+1}^{S_k} p_{ij}\right)\right\} \tag{1}$$

where

$$S_\ell = \sum_{k=1}^{\ell} S_k \quad (S_r = n, \ S_o = 0).$$

If all the components have equal reliability ($p_{ij} = p$), and the matrix is split equally into r submatrices each of m rows and n/r columns, then (1) becomes

$$P = [1 - (1 - p^{\frac{n}{r}})^m]^r \tag{2}$$

for two extreme methods of substitution: the first is the system standby $|r = 1|$ and the second is the element standby $|r = n|$.

If the reliability P is specified the required reliability for each component can be computed from (2).

$$p = [1 - (1 - P^{\frac{1}{r}})^{\frac{1}{2}}]^{\frac{r}{n}}. \tag{3}$$

It is seen from (3) that when the required reliability of a system is close to 1, the component reliability should be at least $\sqrt{n}$ times stronger in case of component substitution scheme than in case of subsystem substitution scheme.

It is assumed that every submatrix has a failure detector, and that whenever one of the operating components fails, the failure is detected and a spare subsystem of components is inserted. P_{ak} - reliability of the detector and switching unit is introduced; then the system reliability equation becomes

$$P = \prod_{k=1}^{r} \left\{ 1 - \left(1 - \prod_{j=S_{k-1}}^{S_k} p_{ij}\right) \prod_{i=2}^{m_k} \left(1 - p_{ak} \prod_{j=S_{k-1}}^{S_k} p_{ij}\right) \right\}. \tag{4}$$

Let the reliability $P_{ij} = P$ be equal for all the components as well as the reliability of the switching units $P_{ak} = P_a$.

A number of operating components in every group (submatrix) is $\frac{n}{r}$ and a number of spare components is $m - 1$. Then (4) becomes (5).

$$P = [1 - (1 - p^{\frac{n}{r}})(1 - p^{\frac{n}{r}})^{m-1}]^r. \tag{5}$$

Comparison of the two redundancy methods (splitting the system to r $l < r$ groups) shows that the number of switching elements in the first case is more than that in the second one. However, system reliability in the first case is always higher. A special case of this statement may be formulated like this: whatever failure detector element reliability may be for any number of operating and spare units, the method of element standby is much more effective than that of system standby.

RELIABILITY vs TIME ANALYSIS

Up to this point, the time interval has been fixed. If it is changed, the reliability becomes a time function. The time moment at which the first failure of the element takes place can be assumed to be a random value ξij and then $1 - P_{ij}(t)$ will be an integral function of this random value distribution. If $P_{ij}(t)$ is continuous the probability density $w_{ji}(t) = -\frac{dp_{ij}(t)}{dt}$ and characteristic function $\delta_{ij}(v) = \int_0^\infty w_{ij}(t)\, e^{ivt}\, dt$ can be derived. "K" order distribution moments will be

$$m_1\left\{\xi_{ij}^k\right\} = \int_0^\infty t^k w_{ij}(t)dt = k\int_0 t^{k-1} p_{ij}(t)dt. \tag{6}$$

From (6), as special case, mean time of failure-free operation is derived,

$$t_{ij}^* = m_1\left\{\xi_{ij}\right\} = \int_0^\infty p_{ij}(t)\, dt \tag{7}$$

as well as time variance of a satisfactory operation.

$$\sigma_{ij}^2 = m_2 - m_1^2 = 2\int_0^\infty t\, p_{ij}(t)dt - \left(\int_0^\infty p_{ij}(t)dt\right)^2. \tag{8}$$

Eqs. (7) and (8) may be obviously used for mean time and time variance of system failure-free operation if we substitute reliability $P(t)$ for $P_{ij}(t)$.

The probability theory technique, given $P_{ij}(t)$, permits us to determine reliability as a time function for system A_{ij} for two limiting methods of switching: hot (all components of a given matrix column are switched in simultaneously) and cold (each component is switched in only if previous one is off).

For reliability under condition of switching in the spare components by the hot method we may obviously use the previously derived equations by substituting constants P_{ij} for function $P_{ij}(t)$.

When switching in the spare components by the cold method the reliability $P_{ij}(t, \xi_{ij-1})$ depends on the random parameter $\xi_{i,j-1}$ —the time instant when the component $A_{i,j-1}$ fails. Thus consider only the mean reliability of component A_{ij}, that is:

$$P_{ij}(t) = m_1\left\{P_{ij}(t, \xi_{i,j-1})\right\},\ (\xi_{i,o} = 0)\ . \tag{9}$$

The characteristic function of the random parameter $\xi_{i,j-1}$ owing to the independence of components becomes

$$\theta_{i,j-1}(v) = \prod_{k=1}^{j-1} \delta_{i,k}(v). \tag{10}$$

Then it is easy to show that

$$P_{ij}(t) = \int_0^t W_{i,j-1}(x)P_{i1}(t-x)dx + \int_t^\infty W_{i,j-1}(x)dx, \tag{11}$$

where $W_{i,j-1}$

$$W_{i,j-1}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \theta_{i,j-1}(v)e^{-ivt}\, dx. \tag{12}$$

REPLACEMENT OF FAULTY ELEMENTS

Further analysis of reliability theory in the way of failure counting is somewhat significant.

Let ξ be a random instant time of failure of a component switched at $t = t_o$. Under the assumption that the failure probability in $(t_o, t_o + t)$ portion of time depends only upon its length and does not depend on the t_o moment of switching, integral function of random variable distribution can be computed from

$$P\left\{t_o \leqslant \xi \leqslant t_o + t\right\} = \int_0^t w(t)dt \tag{13}$$

where $w(t)$ is the random variable distribution density.

Let us name the portion of the time that system will be out of operation—a down time. The down time is a summation of a number of random values: a random portion of time required for searching a faulty element (η_1), a random portion of time required for repair (η_2), a random portion of time required for maintenance (η_3) and so on. If $\eta_1, \eta_2, \eta_3 \ldots$ are independent, a characteristic function method is suitable for down time distribution analysis.

We can assume failure search time initiated at t_1, repair time initiated at t_2, and maintenance time initiated at t_3 are independent of t_1, t_2 and t_3 respectively. Their probability depends however upon specified down time.

Let $\theta\eta_1(v)$, $\theta\eta_2(v)$, $\theta\eta_3(v)$ be characteristic functions of random values η_1, η_2, η_3. Then the characteristic function $\theta\eta(v)$ of their summation $\eta = \sum_i \eta_i$ (that is of down time) becomes

$$\theta_\eta(v) = \prod_i \theta_{\eta_i}(v) \,. \tag{14}$$

Let us also assume that time of satisfactory operation of the element and down time are independent random values. The sum of these random values $\zeta = \xi + \eta$ determines a random failure separation. Let us introduce the characteristic function of a random value ξ

$$\theta_\xi(v) = \int_0^\infty w(t)e^{ivt}\,dt. \tag{15}$$

The characteristic function $\theta(v)$ of the random value (ξ) is determined by the product

$$\theta_1(v) = \theta_\xi(v)\,\theta_\eta(v). \tag{16}$$

The calculation can now be made of probability $F_1(t)$ that failure separation does not exceed t

$$F_1(t) = \frac{1}{2\pi}\int_0^t\int_{-\infty}^\infty \theta_1(v)e^{-ivx}\,dv\,dx \tag{17}$$

where

$$W_n(t) = \frac{1}{2\pi}\int_{-\infty}^\infty [\theta_1(v)]^n e^{-ivt}\,dv. \tag{18}$$

Let $\phi_r(t)$ be probability of n failures in $(0,t)$ portion of the time. Then summing of the probabilities gives

$$\phi_n(t) = F_n(t) - F_{n+1}(t). \tag{19}$$

Under the assumption the number of failures in $0,t$ portion of the time is a random value ν, the probability of $\nu \leqslant n$ can be computed from

$$P\left\{\nu \leqslant n\right\} = \sum_{r=0}^n \phi_r(t) = 1 - F_{n+1}(t). \tag{20}$$

This equation determines probability of satisfactory operation of the given element by providing "n" spare elements in $(0,t)$ portion of the time.

Let us derive the expression for $M(t)$, the mean of the statistical random variable ν, (i.e., the mean failure rate in time t), by introducing the Laplace transformation of function $W_1(t)$

$$\theta_1(ip) = \int_0^\infty W_1(t)e^{-pt}\,dt, \tag{21}$$

where (p) is a complex variable.

Thus

$$M(t) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \frac{\theta_1(ip)}{1-\theta_1(ip)}\, e^{pt}\,\frac{dp}{p}\,. \tag{22}$$

We can also express $W_1(t)$ via $M(t)$

$$W_1(t) = \int_{a-i\infty}^{a+i\infty} \frac{p\mu(p)}{1+p\mu(p)}\, e^{pt}\,dp, \tag{23}$$

where $\mu(p) = \int_0^\infty M(t)\,e^{-pt}\,dt.$

Using the expression (23), we can compute the time between failures probability density $W_1(t)$ for any given mean statistical number of failures for any given (specified) time t. The average number of failures referred to interval of time t, $\left(\lambda(t) = \frac{M(t)}{t}\right)$ according with the established terminology in the probability theory, may be defined as the failure rate.

It is obvious that the value of $\lambda(t)$ is the *a priori* failure probability in time interval $(t, t + dt)$. The asymptotic value of the failure rate in case when "t" is too high can also be easily found.

Since $M(t) = 0$ when $t < 0$, by using the Tauber theorem for the Unilateral Laplace transformation, the following equation can be obtained:

$$\lim_{t\to\infty} \lambda(t) = \lim_{p\to 0} p^2 \frac{\theta_1(ip)}{p[1-\theta_1(ip)]} = \frac{1}{\tau^*}\,, \tag{24}$$

where τ^* is the mean time interval between the two successive switchings on of an element.

If the average failure-free operating time of an element is t^* and the average down time is t_3^* then, due to their independence,

$$\tau^* = t^* + t_3^*\,.$$

Let us consider the probability of the "K" element $(K = 2, \ldots r)$ failure in time interval $(t, t + dt)$ provided that the failure of the basis element occurred when $t = 0$.

Using the rule of summing probabilities one can obtain

$$P\left\{t \leqslant z^{(1)} \leqslant t + dt \; \ldots \qquad t \leqslant z^{(2)} \leqslant t + dt \ldots\right\} = \sum_{n=1}^{\infty} W_n(t)dt \tag{25}$$

$$m(t) = \frac{dM(t)}{dt} = \frac{P\{t \leqslant z^{(1)} \leqslant t + dt \ldots \; t \leqslant z^{(2)} \leqslant t + dt \ldots\}}{dt}.$$

Thus the value $m(t)dt$ is the *á posteriori* probability of failure in time interval $(t, t + dt)$.

It should be taken into consideration that the function $m(t)\,dt$ is not the probability density of a random value and that consequently it does not have the elementary features of distribution functions (for example $\int_0^\infty m(t)dt$ diverges).

The function $m(t)$ may be defined as the failure differential probability of an element at a given time moment regardless of failure rate before that moment. From (22) and (25) one can readily find

$$M(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{\theta_1(ip)}{1 - \theta_1(ip)} e^{pt} dp. \tag{26}$$

Using the above procedure we can obtain the asymptotic value of the failure differential probability of an element after termination of a long time interval since the failure of the basic element

$$\lim_{t \to \infty} m(t) = \lim_{p \to 0} p \frac{\theta_1(ip)}{1 - \theta_1(ip)} = \lim_{t \to \infty} \lambda(t) = \frac{1}{\tau^*}.$$

Let us consider some typical cases illustrating the theory.

Instantaneous Switching of a Spare Element

In case of an instantaneous switching of a spare element the probability density of the random value η is a delta function and the distribution of time intervals between failures coincides with the distribution of the failure-free operating time of the element. For the exponential element reliability distribution law if the average time is t^* we can obtain

$$\phi_n(t) = \frac{t^*}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{dp}{(1 + pt^*)^n} = \frac{1}{n!} \left(\frac{t}{t^*}\right)^n e^{-t/t^*}. \tag{27}$$

This expression is known as the Poisson distribution law, and its presence in the given case might have been predicted.

The probability of no more than "n" failure counts during time t is:

$$P\left\{\nu \leqslant n\right\} = \sum_{r=0}^{n} \frac{1}{r!}\left(\frac{t}{t^*}\right)^r e^{-t/t^*} = \frac{1}{n!} \int_{t/t^*}^{\infty} Z^n e^{-z} dz$$
$$= 1 - \frac{\Gamma(n + 1, t/t^*)}{\Gamma(n + 1)}, \tag{28}$$

where $\Gamma(n + 1, t/t^*)$ is an incomplete gamma function.

The failure rate $\lambda(t)$ coincides in this case with the differential probability of the switching of an element and is constant for any time moments and equal to $1/t^*$.

The reverse relationship can also be readily proved: if the failure rate is constant for any t the distribution law of time intervals between failures is also exponential.

Taking the formula of the operational calculus it can be shown that

$$\lambda(t) \to \int_p^\infty \mu(s)dS$$

and if $\lambda(t) = \frac{1}{t^*} = \text{const}$, then $\int_p^\infty \mu(s)dS = \frac{1}{pt^*}$, (29)

and thus

$$\mu(p) = \frac{1}{p^2 t^*}$$

or

$$W_1(t) = \frac{1}{t^*} e^{-t/t^*}.$$

It can be seen from these equations that the failure rate will be constant if the distribution law of time intervals between failures is exponential.

Constant Down Time

Let us assume that down time is constant and is equal to t_3^*, and the reliability distribution law is exponential. Under these assumptions the probability density of the failure duration is $\delta(t - t_3^*)$.

The probability of exactly "n" failure counts during time interval $(0 - t)$ is

$$\phi_n(t) = \frac{\Gamma\left(n, \frac{t - nt_3^*}{t^*}\right)}{\Gamma(n)} - \frac{\Gamma\left(n+1, \frac{t - (n+1)t_3^*}{t^*}\right)}{\Gamma(n + 1)},$$
$$t \geqslant (n + 1)\, t_3^*. \tag{30}$$

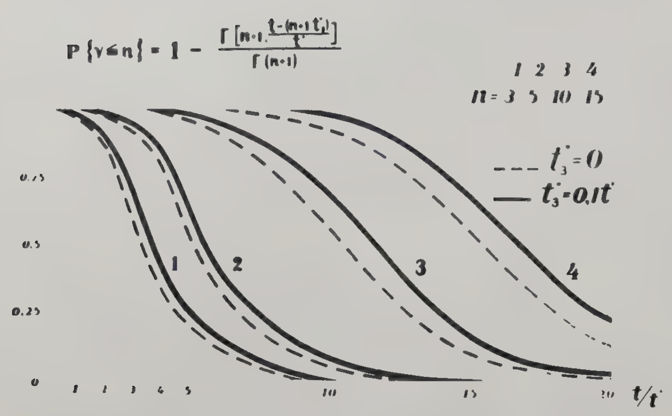


Fig. 1—Probability of not more than n failures in time t.

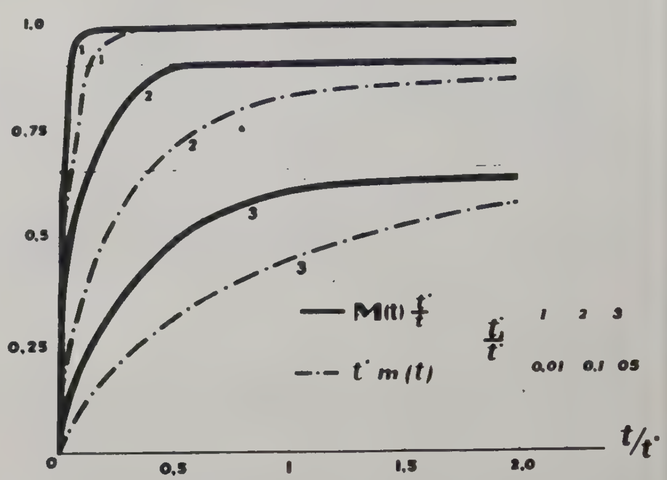


Fig. 2—Failure rate and differential failure probability. Exponentially distributed down time.

The probability of failure counts within time interval t not exceeding "n" is:

$$P\{\nu \leqslant n\} = 1 - \frac{\Gamma\left(n+1, \frac{t-(n+1)t_3^*}{t^*}\right)}{\Gamma(n+1)},$$

$$t \geqslant (n+1)\, t_3^*. \tag{31}$$

If $n > 30$ we may take advantage of the asymptotic equation

$$P\{\nu < n\} \sim 1 - F\left(2\sqrt{\frac{t-(n+1)t_3^*}{t^*}} - 2\sqrt{n+1}\right), \tag{32}$$

where F is a well tabulated Laplace function. Fig. 1 shows the curves of probability $P\{\nu \leqslant n\}$ vs t/t^* for various "n" and $t_3^*/t^* = 0.1$. The dotted line corresponds to the probability for the limit case when $t_3^* = 0$, that is for the Poisson distribution law.

The Down Time Exponential Distribution Law

In conclusion it would be of some interest to examine the case when the failure duration as well as the operating time remains unchanged in distribution by the exponential distribution law. In this case if $t_3^* \ll t^*$

$$\phi_n(t) = \frac{1}{n!(n-1)!}\, e^{-t/t^*} \int_0^{t/t^*} y^{n-i}\left(\frac{t}{t^*} - y\frac{t_3^*}{t^*}\right)^n e^{-\left(1+\frac{t_3^*}{t^*}\right)y}\, dy. \tag{33}$$

The failure rate is

$$\lambda(t) = \frac{1}{t^*+t_3^*}\left[1 - \frac{t^* t_3^*}{(t^*+t_3^*)t}\left(1 - e^{-\frac{(t^*+t_3^*)t}{t^* t_3^*}}\right)\right], \tag{34}$$

and the differential failure probability in time t after failure of the basic element is

$$m(t) = \frac{1}{t^*+t_3^*}\left[1 - e^{-\frac{(t^*+t_3^*)t}{t^* t_3^*}}\right]. \tag{35}$$

The curves of the failure rate and differential failure probability vs t_3^*/t^* are shown in Fig. 2.

PURCHASING RELIABILITY*

E. J. BREIDING†

Summary—The contributing responsibility and a system for controlling component reliability as part of the procurement function are discussed. The system, which is particularly applicable to large-scale procurement, consists of three major controls:

1) Component Vendor Approval Procedure—a tool for obtaining and documenting bona fide sources capable of supplying components to specified time and quality requirements.
2) Vendor Delivery Performance—a purchasing control for maintaining and upgrading performance from selected vendors.
3) Vendor Quality Rating—a quality control tool to aid purchasing in maintaining and upgrading quality of products from selected vendors.

COMPONENT VENDOR APPROVAL PROCEDURE

The first step for obtaining *bona fide* sources for components is to establish an approved component list. This is initiated at the earliest possible engineering stage. Vendors are approved for a development list after it has been established by specification and performance analysis and, then, by initial survey that they can satisfy the requirements of development and/or product engineering, quality control, and purchasing.

Actual writing of the component specification is the responsibility of the component application engineer who normally supplies purchasing with at least one producing source for the component. During the cycle of releasing a system, engineering may aid by making purchasing aware of the technical state of the art so that they can determine the names and numbers of sources required for a given component.

It is, therefore, particularly important that the purchasing department be qualified to establish the proper criteria for multiple-sourcing of components on the list.

The following criteria are suggested as the basis for a formal evaluation program to equip purchasing with data on which to advise priorities for sourcing evaluation.

1) Cost of the component or its type per system. Estimated quantity to be used must be known so that a descending dollar value can be established.
2) Degree of standardization or lack of standardization resulting from:
   a) technical state of the art,
   b) proprietary rights.
3) History of existing vendors' delivery performance.
4) History of existing vendors' quality performance.
5) Comparative failure analyses based on field failure data. Such data may be difficult to obtain due to the time required to make a significant survey; however, it is advisable to initiate this program where practical.
6) Obviously, other factors, such as tooling cost and time, must be considered.

When the need for a component in production quantity is definitely established, suggested vendors are submitted for approval to product engineering, quality control, and the engineering buyer. Approvals are secured as illustrated in the flow chart in Fig. 1.

In relation to the SAGE program, manufacturing engineering reviews the suggested vendor's products and shipping containers to determine if they are adaptable to the high-speed production essential to SAGE manufacturing operations. If the product is practical for this mechanization, or can be brought to standards by the vendor, approval is granted by manufacturing engineering.

Accurate assessment by purchasing of the vendor's production facilities is a critical factor in determining the product quality and delivery performance which can be safely anticipated from a given source. Facility survey is a function of quality control, and its inspection should include: quality, inspection, calibration, and parts and record control techniques.

*This paper was presented at Thirtieth Radio Fall Meeting, Rochester, N. Y., October 27-29, 1958.

†International Business Machines Corp., Kingston, N. Y.

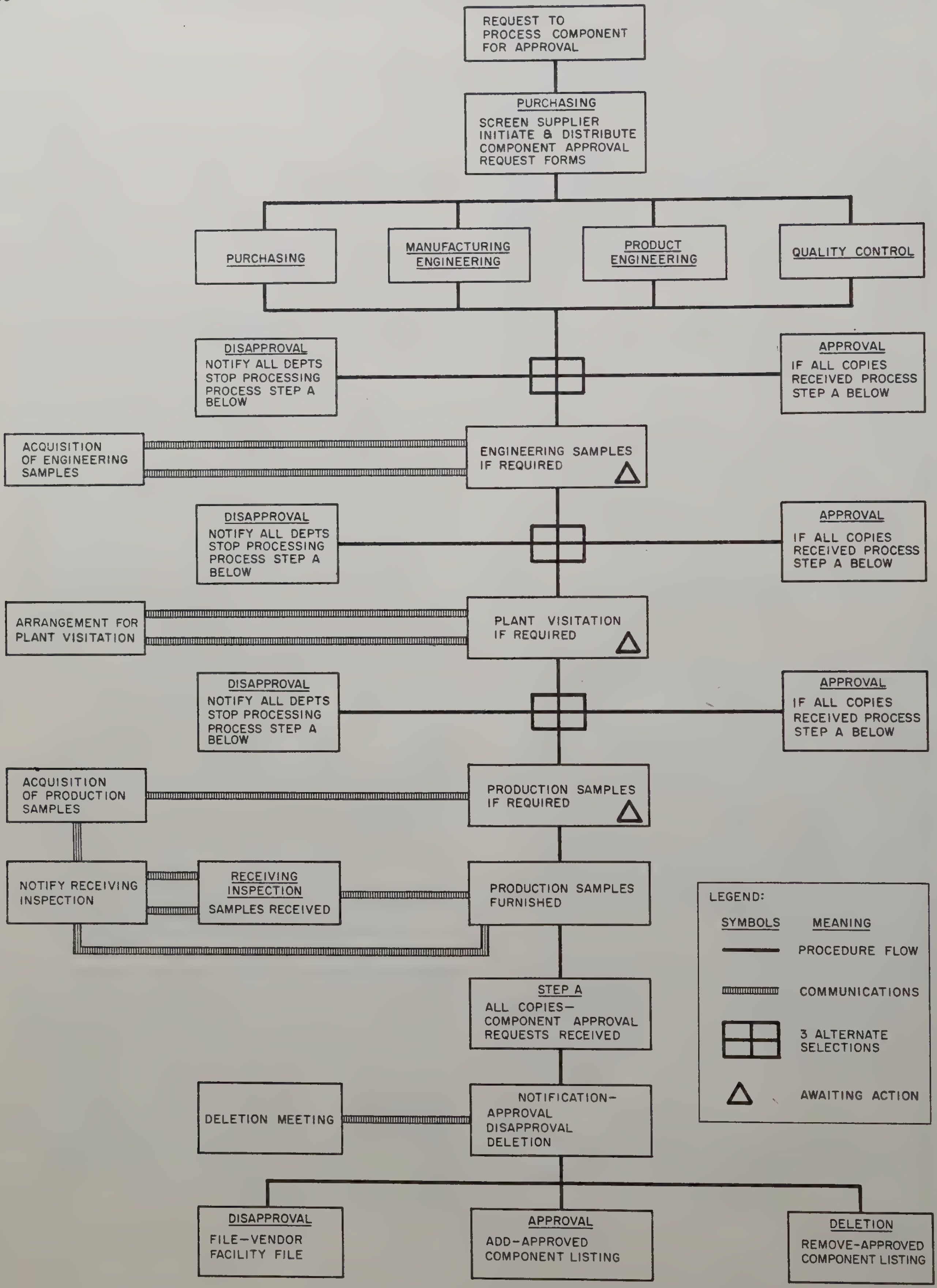


Fig. 1 —Flow chart component approval.

VENDOR DELIVERY RATING

Approved vendors are listed on individual vendor record cards (see Fig. 2). Both the vendor delivery and the quality ratings are graphically presented on the card and are compatible in that

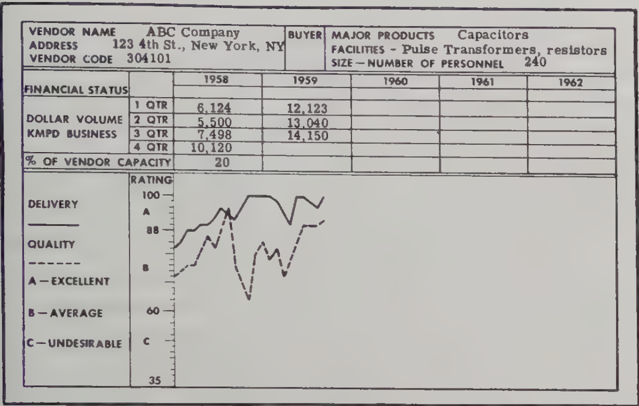


Fig. 2 —Vendor record card.

both use the 0 to 100 scale with identical cut-off points for excellent, average, and undesirable. A ready reference is now available which the buyer uses to review quality and delivery performance of a particular vendor.

The delivery rating system utilized for these evaluations was developed and initiated at IBM Military Products Division, Kingston, N. Y., to deal with the great number of vendors involved in purchasing for the SAGE computer system.

The rating is determined by adding the total number of days late on each delivery and dividing by the total number of shipments. This gives the average number of days late. The resultant rating is then classified according to a scale ranging from 0 to 100, with 88 to 100 defined as excellent, 60 to 88 as average, and below 60 as undesirable. The method provides a formula which impartially evaluates each vendor, whether large or small and frequently or infrequently used, and graphically displays the rating so that trends and patterns can be immediately discerned.

Significantly, with use of the control, the overall average monthly delivery rating jumped to 91.7 from an initial statistical average of 75.

VENDOR QUALITY RATING

As Fig. 2, indicates, quality of performance is also charted on the vendor record card.

The quality rating is based on a statistical average rating of 75 which occurs when the sample fraction defectives are equal to the specified AQL's, with a maximum of 1 in 25 lots rejected. A rating of 60 occurs when the sample fraction defectives equal 0.5 times the specified AQL's, with a maximum of 1 in 156 lots rejected.

Complete information on the procedure is available in a brochure published by the Kingston Military Products Division of IBM, Kingston, N. Y.[1]

As Table I indicates, the rating method itself equalizes rating inequalities common to some other rating systems. The visual method of displaying current status makes the chart understandable and practical for both vendor and buyer reference. The buyer has the added advantage of being equipped with a quick and accurate visual history which immediately pinpoints any vendor trends which could affect supply or quality. The system has been markedly beneficial in establishing a smoother flow of components from the vendor and in preventing production crises that interrupt the procurement of products of required reliability.

VENDOR SUPPORT

It is advisable that vendors be fully acquainted with the philosophy of the rating systems to realize maximum effectiveness from the program. The IBM Kingston Procurement and Quality Control departments jointly sponsored a program to acquaint 150 of their suppliers with the purpose of the controls and methods used for measuring their performance. Engineering-type seminars and demonstrations are held on a regular basis to assist the vendor in satisfying quality and performance requirements and to improve vendor products and relations.

OTHER RELIABILITY DATA

Usually, the responsibility for component reliability is associated with engineering, as it concerns circuit design, component evaluation, and maintenance techniques. Numerous reports covering these engineering aspects of reliability have been made, including a trilogy entitled, "Reliability of an Air Defense Computing

[1]H. G. Harding and J. Rowinski, "Vendor Quality Rating Procedure, IBM Corp., Kingston, N.Y., Brochure; March 15, 1957.

TABLE I

COMPARISON CHART

Factor	Other Rating Systems	KMPD Rating System
Rating of 100	Rating of 100 can be obtained if defectives are present in sample and sometimes may not be obtained even if no defectives are found.	Rating of 100 is obtained only if no defectives are found in sample.
Effect of quantities submitted upon rating	Vendors submitting small lots may deserve, and not obtain, high ratings because ratings are affected by lot size.	Vendors submitting lots with few defectives are given high ratings regardless of quantity.
Standard deviations	Standard deviations are accepted as criteria and involve lot size.	ACL values, to which the vendor adjusts the quality of his material, are used in place of standard deviations.
Rating limits	Rating limits are relative to standard deviations, instead of being specific values known to the vendor.	Rating limits are based on 0.5 and 1.6 AQL's and are easily understandable to the vendor.
Mathematics required	Calculations involve a combination of formulas and complex procedures.	Calculations involve only sample results and a corresponding constant for each AQL, as needed.
Interpretation of ratings	Ratings are influenced by factors which the vendor cannot control. They cannot be used by the vendor as an accurate guide for improving his product.	Ratings reflect only the actual quality of the product. The vendor can rely on accuracy of the ratings.

System."[2] The Delivery Performance System described in this paper has also been reviewed in the literature.[3]

[2] H. F. Heath, Jr., "Reliability of an air defense computing system: component development," IRE TRANS. ON ELECTRONIC COMPUTERS, vol. EC-5, pp. 224-226; December, 1956.
R. E. Nienburg, "Reliability of an air defense computing system: circuit design," IRE TRANS. ON ELECTRONIC COMPUTERS, vol. EC-5, pp. 227-233; December, 1956.
M. M. Astrahan and L. R. Walters, "Reliability of an air defense computing system: marginal checking and maintenance programming," IRE TRANS. ON ELECTRONIC COMPUTERS, vol. EC-5, pp. 233-237; December, 1956.

[3] "Rating system pinpoints delivery and quality trends," Purchasing News, July 28, 1958.

CONCLUSIONS

Procurement Departments must, in the future, make full use of measurement and control techniques to assure that their selection of vendors has assisted the over-all reliability program. The control system described in this paper has demonstrated several distinct advantages in operation that recommend its use.

DIAGNOSIS OF EQUIPMENT FAILURES*

J. D. BRULÉ,† R. A. JOHNSON,† and E. J. KLETSKY†

Summary—This paper introduces several new concepts which are applicable to the problem of diagnosis of equipment failures. Following the definitions of an equipment, an element of the equipment, and the model of a test, a general diagram of a testing procedure is developed. The testing diagram is constructed in such a way that the various tests needed and the probability of failure of the elements are readily incorporated. While it is found that a completely general testing diagram becomes quite complicated even when the equipment under consideration is not intricate, a major simplification is obtained by introducing a simplified diagram with suitably restricted tests. This simplified testing diagram may be used repeatedly in order to find all the faulty elements of the equipment.

With reference to the testing diagram, it is possible to compute the minimum average cost of diagnosing the equipment. This appears to be the most useful measure of the efficiency of a test procedure. The order of magnitude of this optimization problem is discussed and solutions for two special cases are obtained by analogy with an optimum coding problem.

INTRODUCTION

Recent advances in the design of electronic systems have resulted in an ever-increasing complexity in such systems. While such complexity is necessary to permit the systems to perform tasks of ever-increasing scope, the problem of keeping the systems in working order tends to increase at least as fast as the basic complexity. The present investigation is concerned with one particular phase of the problem of maintenance—namely, the problem of determining which part of an equipment is in need of repair when the equipment as a whole does not function properly. While this is only one aspect of the whole problem of maintaining equipment in working order, any improvement in general diagnostic procedures would result in significant economies.

Some of the other factors which contribute to maintainability are the reliability of the basic component parts, proper preventive maintenance procedures and proper mechanical and electrical design so as to make component parts easily accessible for replacement. Despite the advances in component reliability and preventive maintenance, it appears to be the "nature of the beast" that a great many unpredictable or random failures do occur. It is for such failures that the importance of efficient diagnostic procedures are important. Equipment logs indicate that a significant part of the "down time" of equipments is spent in diagnosing what part of the equipment does not function properly. Such records also show that incorrect diagnoses occur frequently. For example, tubes which have been replaced as defective are often found to be perfectly acceptable.

The present work was undertaken in the hope that examination of the basic fundamentals of diagnostic procedures would permit substantial economies to be made in the diagnostic phase of maintenance. The need for systematized procedures is increased by the fact that in many cases equipment must be maintained by semiskilled technicians whose training does not enable them to understand completely the functioning of the complicated machines which they are to maintain. We do not propose that any results we obtain would necessarily aid a highly skilled technician with adequate experience on the equipment in question. However, when the training or experience is inadequate, systematized procedures should result in economies. Another factor which favors systematized diagnostic procedures is the present trend towards automatic equipments and, in particular, equipments which are capable of diagnosing failures of their own component parts. It is axiomatic that any function which is to be performed by a machine must first be systematized.

A review of the technical literature yields a surprisingly small number of references to the general problem of efficient or optimum diagnostic procedures. A recent paper by Hoehn and Saltz [1] which summarizes previous work gives only two references [2], [3]. Hoehn and Saltz discuss two approaches to efficient diagnosis. Diagnostic procedures which are applicable to specific equipments have been published (usually in the form of maintenance manuals, many of

*This work was partially supported by Rome Air Dev. Center under contract AF 30(602)-1833.
†Electrical Engrg. Dept. Syracuse University, Syracuse, N. Y.

which are classified) but the basic aspects of the diagnostic problem are then hidden by the constraints of the specific equipment. A recent paper [4] on the identification of the various types of malfunctions of a syncro-repeater system is an example of this.

A problem related to the diagnosis of equipment failures is the initial check-out or "debugging" of equipments. Although this problem is not considered in detail here, many of the techniques developed for diagnosis are also applicable to check-out.

MATHEMATICAL MODELS

In order to apply mathematical techniques to the problems encountered in the testing of equipment, it is necessary to formulate general mathematical models of an equipment and a test procedure. If such models are to be applicable to a large class of equipments, they must necessarily be somewhat abstract. Accordingly, one should not expect that the models contain all the details peculiar to a given equipment or a given test procedure. However, the models must be capable of representing the essential features and organization of virtually any equipment and the various procedures by which it can be tested.

Model of an Equipment

As a start toward developing a model of an equipment, the following definitions are made:

Equipment: For test purposes, an equipment is defined as a collection of functional elements which are interconnected so as to generate specified responses on the application of a specified set of primary stimulants. The primary stimulants are independent of the operation of the equipment.

Functional Element: A functional element is defined as a constituent part of an equipment which generates a single response on the application of a specified combination of stimulants which may include the responses generated by other functional elements.[1]

These two definitions implicitly define a convenient diagram of an equipment which is to be the subject of a testing procedure. An example

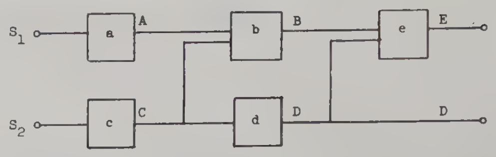


Fig. 1 —Diagram of an equipment.

of such a diagram is shown in Fig. 1. In this diagram the functional elements are denoted by lower case letters and the responses of the elements by the corresponding upper case letters. S_1 and S_2 represent the primary stimulants to the equipment as well as the stimulants to elements a and c, respectively. E and D represent the responses of the equipment as well as the responses of elements e and d, respectively. For a complete specification of the equipment, it is necessary to specify which combinations of stimulants are required for each element to generate the specified output. In the simplest case, the specified response is generated only when all of the stimulants to that element are present and the element in question is functioning properly. Inspection of Fig. 1 shows that with this assumption the specified equipment responses (E and D) will not be present unless S_1 and S_2 are present and all of the elements are functioning properly.

The definitions of equipment and, particularly, functional element given above are not sufficiently detailed to determine completely the description of an equipment for test purposes. This freedom in the designation of the elements is necessary if the model is to be useful in representing actual diagnosis procedures. For example, in a military situation, a large system may include radars, computers, and guns. The whole system may be described to be the "equipment" and the radars, computers, and guns to be the "functional elements." This description would be appropriate for the personnel responsible for the actual control of the system. For such personnel the diagnosis of a malfunction of the system is merely the determination of which radar, computer, or gun is not functioning properly. If it is found that the radar is at fault, the responsibility for further diagnosis is shifted, for example, to a radar technician. To the technician the radar is the equipment rather than the functional element and the transmitter, receiver, sweep generators, etc., are the functional elements. Having located the fault, say, to the receiver, the receiver becomes the equipment and the IF amplifier, local oscillator, video

[1]In common engineering terminology the responses are "outputs" and the "stimulants" are the "inputs." The terminology used is intended to be as general as possible.

amplifier, etc., become functional elements. Thus the diagnosis proceeds through a number of levels. The model of an equipment that is introduced is applicable to each level by proper designation of "equipment," and "functional element," and the diagram of an equipment at each level follows when the functional elements have been identified and the stimulants and responses specified.

Two possible sources of complication which are inherent in the definitions of equipment and functional element have not been included in the simple example of Fig. 1. In the first place, there are no loops in the diagram such that any response depends on itself through feedback. An equipment with feedback may be tested by breaking the loop and treating the response at the break as an additional response and primary stimulant.

The second complication which may arise in the diagram of an equipment is the presence of redundancy. This problem is discussed briefly by Johnson, *et al.* [7], where alternative means for testing redundant systems are presented. One can determine if all paths are functioning by breaking all the redundant paths and replacing them in turn or by designating the output of each redundant path as an equipment response. Alternately, one might add an element to the equipment which would have an output only when all redundant paths are functioning.

Model of a Test

The diagram of an equipment introduced above implicitly defines the tests which can be performed in diagnosing the equipment for the purpose of determining which element, if any, is not functioning properly. Such tests consist of supplying stimulants to the elements and observing the responses. For example, (with reference to Fig. 1) if stimulants A and C are applied to element b and the response at B meets the specifications, element b is known to be functioning properly. Similarly, if S_1 and S_2 are applied and response B satisfies the specifications, we can infer that elements a, b and c are functioning properly.

In order to avoid repetition of the phase "functioning properly," the following terms shall be used in the report:

"Good" element: an element which functions properly. When the specified stimulants are supplied to the element, the specified response from the element is developed.

"Bad" element: an element which does not function properly. When the specified stimulants to the element are supplied, the specified response from the element is not developed.

"Questionable" element: an element which is not known to be good.

In these terms, a diagnosis test procedure consists of identifying the bad elements, if any, in a set of questionable elements. Similarly a check-out test procedure consists of determining whether all elements are good which, in turn, assures that the equipment is functioning properly. In most test procedures the determination of which questionable elements are, in fact, good or bad is made by logical deduction from the results of a number of tests.

Any given test will pass if the elements being tested are all good, and will fail if one or more of these elements are bad. For an equipment with N elements, a possible representation (and also designation) of a test is a sequence of N symbols, one symbol for each element. A given test examines a subset of k of these elements to determine if they are good and ignores the remaining N - k elements. The symbol 0 is assigned in the jth position of a test if the jth element of an equipment must be good in order for the test to pass. The symbol 1 will be placed in the jth position of a test if the jth element is not tested.

This notation makes it possible to construct a complete list of all the tests which can be performed on an equipment for which an equipment diagram is available. Table I contains such a list for the example shown in Fig. 1. All of the tests are performed by applying a set of stimulants as listed in the second column and observing the response indicated in the first column. The third column lists the elements which must be good if the specified response is to be observed. The numerical designation of the test is given in the last column. The digits of the test designation are 0 if the corresponding element must be good in order for the response to be within specifications (*i.e.*, for the test to pass). The total number of different tests which can be defined for an equipment with N elements is obviously equal to the number of N-digit binary numbers. However, the test which has a numerical designation 111. . . . . 1 is of no significance since the result is independent of whether the elements are good or bad. Accordingly, the total number of significant tests is given by

$$2^N - 1 .$$

For the example (Fig. 1) only the 19 tests listed in Table I are realizable in accordance with the

TABLE I

Tests Associated with the Equipment of Fig. 1

Response Observed	Stimulants Required	Good Elements to Pass	Numerical Designation
			abcde
A	S_1	a	01111
B	AC	b	10111
C	S_2	c	11011
D	C	d	11101
E	BD	e	11110
B	CS_1	ab	00111
B	AS_2	bc	10011
B	S_1S_2	abc	00011
D	S_2	cd	11001
E	ACD	be	10110
E	CDS_1	abe	00110
E	ADS_2	bce	10010
E	DS_1S_2	abce	00010
E	BC	de	11100
E	BS_2	cde	11000
E	AC	bde	10100
E	CS_1	abde	00100
E	AS_2	bcde	10000
E	S_1S_2	abcde	00000

definition of a test. The remainder of the 31 tests can only be realized as combinations of these 19 tests. For example, the result that the tests 11001 and 10011 both passed (or both failed) is equivalent to knowing the result of test 10001. In either case the conclusion is that elements b, c, and d are all good (or at least one has failed).

A somewhat different type of test, not necessarily included in the above tabulation, is a test that is made by replacement. In a replacement test, a questionable element is replaced by an element that is known to be good, various stimulants are applied and a response is observed. Two such tests are listed below for the example of Fig. 1.

Replace Element	Observed Response	Stimulants Required	Good Elements to Pass	Numerical Designation
c	E	S_1S_2	abde	00100
b	E	S_1S_2	acde	01000

The test 00100 is equivalent to one on the list, while the test 01000 is different. However, this example shows that the proposed designation of a test is capable of representing replacement tests.

As a final comment, note that the first five tests listed in Table I are single element tests in that they have but one 0 in their designation. As such, failure to pass one of these tests provides unambiguous information that the corresponding element is bad.

Testing Procedures

Testing procedures may be organized in two essentially different ways. In the first, a number of tests are performed and the results analyzed to determine which elements are good and which are bad. Note that in this case the order of performing the tests is immaterial since all the tests are done before analysis is attempted. Such a procedure will be termed "combinational" in that the analysis depends on the combination of the results of the tests. If it is assumed *a priori* that only one element or a small fraction of the elements are bad, such a procedure is inefficient in that it requires that all the tests be done in every testing procedure. In an alternative procedure, the choice of the next test to be done is dependent on the results of the previous tests. Such a procedure will be called "sequential" in that the analysis of the test results is carried out sequentially. Most diagnostic procedures are sequential in nature and essentially consist of localizing a bad element. Check-out (check-up in medical terminology) testing procedures, however, are not sequential in that a number of tests are required and the order of performing the tests is not important. The representation of a test procedure developed below is directly applicable to the sequential procedure since our primary interest is in diagnosis.

A convenient method of stating the cumulative information[2] gain from the results of several tests is to define "information states" in which the status of each element is given as good, bad, or questionable. This concept enables one to draw a diagram of a sequential testing procedure in which new information states are generated from previous states and the results of intervening tests. The basic building block in such a

[2]Here, "information" is used in the qualitative sense. A quantitative development of this concept is presented in a companion paper by R. A. Johnson, "An information theory approach to diagnosis," to be presented at Sixth National Symposium on Reliability and Quality Control, Washington, D. C., January 11-13, 1960. Abstract, this issue, p. 35.

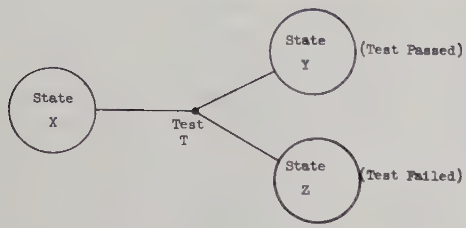


Fig. 2 —Basic building block of a sequential test diagram.

diagram is shown in Fig. 2. In general, a complete testing procedure results in a single final state in which the status of each of the N elements is known to be either good or bad. Since such a state can be specified by an N-digit binary number, it is evident that there are 2^N final states possible. The number of building blocks (Fig. 2) and, consequently, the number of tests which appear in a complete diagram is $2N - 1$. Thus, even for a small number of elements, the diagram of a complete sequential test procedure becomes quite complicated.

A Simplified Sequential Test Procedure

A considerable simplification of the diagram of a diagnostic testing procedure is obtained if the assumption is made "that one and only one element is bad." Such a diagram will be referred to as a simplified sequential test diagram. In this diagram the number of final states is equal to the number of elements N since any one element may be bad. Also, the diagram has on it exactly $N - 1$ test vertices. This is seen by noting that at any state in the testing diagram where there are N questionable elements, a single test will divide the N elements into two subgroups of N_1 and N_2 elements, where $N = N_1 + N_2$. For a test to be nontrivial, both N_1 and N_2 must be equal to or greater than one. This property of a test can be shown quite simply by representing the N elements as a sequence of N symbols on a line, as shown in Fig. 3. In this figure, test T_1 divides the $N = 5$ elements into two subgroups of $N_1 = 2$ and $N_2 = 3$ elements. These two subgroups must each be subdivided until, as stated above, the original group of N elements is

$$\begin{array}{cc|ccc} & & T_1 & & \\ e & e & e & e & e \end{array}$$

Fig. 3 —Symbolic representation of a test.

divided into N subgroups. This is accomplished when a line representing a test is placed between each symbol representing an element. There are $N - 1$ intervals between the N symbols; thus $N - 1$ tests must appear on the testing diagram. Note that this simplified sequential test diagram can be used over and over to find each of the bad elements provided that the presence of more than one bad element does not make it impossible to find at least one.

Referring back to the equipment defined by Fig. 1 and the notation previously introduced, the simplified diagram of a sequential diagnostic test procedure is shown in Fig. 4. Here the initial state has been designated 11111 since at this state each of the elements is questionable. If the first test is passed, elements c and d are known to be good and therefore this state is designated 11001. If the test fails, c or d is bad

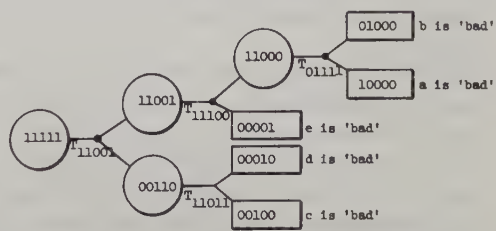


Fig. 4 —A simplified sequential test diagram.

and, by the assumption that only one element is bad, a, b and e must be assumed good. Therefore, the state designation becomes 00110.

A test following state 11001 may be any test which separates elements a, b, and e. The tests T_{11100}, T_{11000}, T_{11010}, and T_{11110} are all equivalent here since c and d are known to be good before the test. From this it is seen that the structure of the testing diagram and assignment of the final states does not uniquely determine the tests which must be used.

Since much of the remainder of the paper is concerned with analysis of simplified sequential testing diagrams of the type illustrated in Fig. 4 a list of the important properties of such diagrams is of interest. Some of these properties are:

1) There is one initial state designated 111. . . 1.
2) There are N final states where N is the number of elements in the equipment. The final states are characterized by a designation which includes one '1' and $N - 1$ '0's.
3) There are a total of $N - 1$ "test

vertices" required independent of the structure of the diagram.

4) The same test (as determined by the test designation) may appear on more than one path from the initial to a final state. (The case of purely combinational testing has the same tests on *all* paths from initial to final states. Thus the model includes combinational testing.)
5) The same test does not appear more than once on the path from the initial state to a *given* final state. (This condition insures that the same test is never repeated in diagnosing a given failure.)
6) The designation of the new state created when a test passes is obtained by multiplying the designation of the previous state and the test designation, digit by digit. By convention this state appears on the upper branch leaving the test vertex.
7) The designation of the new state created when a test fails is obtained by multiplying the designation of the previous state and the complement of the test designation, digit by digit. By convention this state appears on the lower branch leaving the test vertex.

An examination of the maximum number of different tests which can exist for N elements for the simplified sequential test procedure will now be made. As noted above, there are a total of 2^N N digit binary numbers, but the test consisting of all 1's is trivial. In addition, the assumption that one element is bad eliminates the test consisting of all 0's. Also, this assumption means that only half of these $2^N - 2$ tests are useful. This is because a test T and its complement T', (0's and 1's interchanged) supply the same information about the state of the equipment. Consequently, the number of useful (essentially different) tests is now $2^{N-1} - 1$.

It must be recognized that, in general, not all of these $2^{N-1} - 1$ tests may be possible. For example, if the equipment consists of a cascade of 4 elements, a single test cannot be performed which will determine if the first and third elements alone are good. Consequently, for a given set of m tests, it is necessary to establish if these tests are adequate to determine the bad element out of the N given elements.

A method for determining the adequacy of a given set of tests is illustrated in the following example. Consider the set of m = 6 tests for N = 5 elements:

$$T_1 = 10010,$$
$$T_2 = 01110,$$
$$T_3 = 00101,$$
$$T_4 = 10110,$$
$$T_5 = 10101,$$
$$T_6 = 11000.$$

The problem is to check these tests to establish if element 2, for example, can be isolated by any sequence of tests. Rewrite each test, or its complement, in such a way that there is a 0 in the second position and then add the 1's in each column as shown below.

$$\begin{aligned} T_1 &= 10010, \\ T_2' &= 10001, \\ T_3 &= 00101, \\ T_4 &= 10110, \\ T_5 &= 10101, \\ T_6' &= \underline{00111} \\ & 40434 \ . \end{aligned}$$

The set is adequate to test element 2 if the 2nd column is the only one having the sum zero. Note that, in this example, it is not necessary to rewrite any of the tests after the third since at this point it is obvious that none of the columns, other than the second, will have a sum of zero. This procedure must then be repeated for each of the elements. It can be seen that this algorithm is a valid method for determining adequacy by noting that it merely replaces the process of forming products to determine states by the process of addition. Thus, this checking procedure provides a rapid means for determining the adequacy of a set of tests when it is known that exactly one element is bad.

An alternative representation of a sequential test procedure in terms of Boolean matrices is given in Johnson, *et al.* [7]. The representation is completely equivalent to the simplified sequential testing diagram described above. As far as can be determined, the matrix representation does not provide any advantage over the testing diagram.

Restriction of the Simplified Test Procedure

Following the testing diagram defined above is a sufficient procedure for locating which element has failed provided the assumption that "only one element is bad" is satisfied. Consider now the restrictions and modifications which must be

introduced if this assumption is not met. First consider the case where no elements are bad. In this case the results of each test will be "pass" and, with the convention indicated in Fig. 2, the upper branch from each test is followed. This path leads to a specific element e_f (element b in the example of Fig. 4) which, under the assumption that one element is bad, is known to be bad. Note that the deduction that this element is bad is a consequence of the *fact* that all the other elements are known to be good as a result of the tests performed and the *assumption* that one element is bad. If this assumption is removed, the status of element e_f is still questionable since none of the preceding tests required that e_f be good in order to pass. Before the remaining ambiguity can be removed, a test must be performed which will pass only when e_f (or e_f and other elements which are known to be good) is good. Thus the possibility that no elements are bad and the machine as a whole is good can be taken care of by a slight modification of the simplified testing diagram which contains the implicit assumption that one element is bad. An illustration is given in Fig. 5. Note that in this case, the test diagram has exactly N test vertices.

A second possible source of difficulty arises when more than one element is bad. Here, it is sufficient to insure that the presence of additional bad elements will not lead to the identification of any of the good elements as bad. If this condition is met, all the bad elements can be identified by repeated application of the same testing procedure after repairing each bad element as it is identified. The restrictions imposed by this condition are established below.

In the assignment of binary numbers to represent the states in the testing diagram, certain of the 0's have been inserted by deduction using the assumption that one element is bad and the failure of certain tests. Actually, the status of the elements corresponding to these 0's is questionable since the tests involved are tests of the remaining elements which are known (from the test results) to contain one bad element. The questionable status of the elements which have been labeled good by use of the assumption will not influence the results of the remaining tests provided that these elements are not required to be good in order that any of the remaining tests pass. In other words, when a failure has been localized to one group of elements, further tests should not involve elements which are not in this group. The restriction on the allowable tests is illustrated in Fig. 5. In this figure the 0's which are assigned by making use of the "one element is bad" assumption are underlined (<u>0</u>) and the restriction on the following tests as indicated by underlining those 1's in the test designation which insures that the 0's will not influence the results of the test. It is evident that the following condition is sufficient to insure that the presence of more than one bad element will not interfere with the determination of one bad element:

> Each test shall contain a 1 in the position corresponding to any 0 of the preceding state which has been determined as the result of the failure of a previous test.

The X's appearing in the test designations of Fig. 5 indicate that the choice of 1 or 0 is completely immaterial since the corresponding elements are known to be good as a result of previous tests. This may be summarized in the statement.

> The digit of the test designation in the position corresponding to any 0 of the preceding state which has been determined as the result of the passing of a previous test is arbitrary.

The number of distinguishable tests is 2^N-1 in this more general case since the assumption that 1 element is bad is no longer made.

Once again, not all of these tests will be possible, and a method is now established for determining if a given set of $M < 2^N-1$ tests is adequate to determine which elements, if any, of a given equipment are bad if the testing diagram is constructed according to the above rules. Such a set of

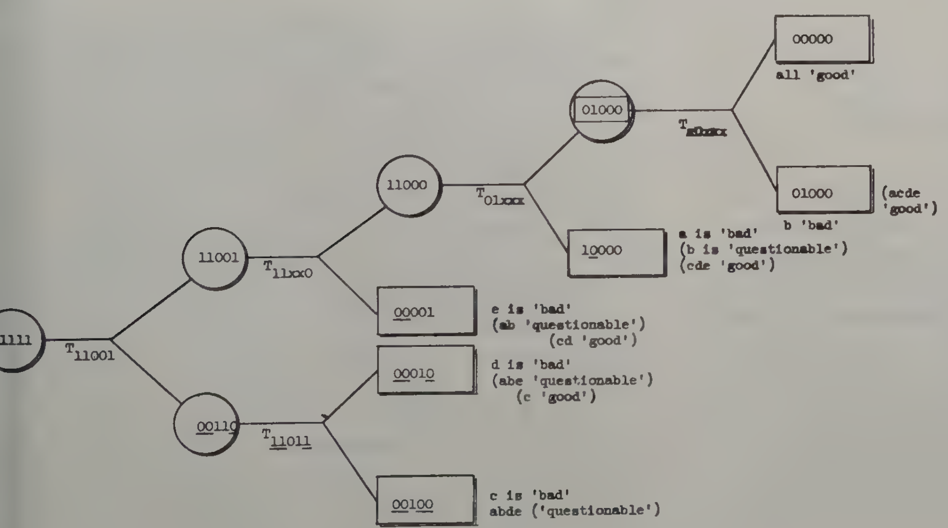


Fig. 5 —Modified simplified testing diagram including restrictions on the tests.

tests will be called "adequate in general."

Considering the situation when each test fails, it is seen that eventually a state consisting of (N-2) 0's and two 1's is reached, and all of the 0's will be underlined. In order to determine which of the two elements are bad, the next test must have 1's in each of the underlined 0 positions of the preceding state and exactly one 1 in one of the remaining two positions. Thus, a necessary condition for a set of tests to be adequate in general is that the set must have at least one test which has (N-1) 1's and one 0. If this conditions is satisfied, then it is possible to isolate this element if it is bad. Hence, we can eliminate this element from each test, and we are left with a modified set of tests for N-1 elements. In order for this modified set to be adequate in general it, too, must contain at least one test which has exactly one 0 in its designation. A repeated application of this check supplies a necessary and sufficient condition for a given set of tests to be adequate in general. For example, consider the set of tests,

$$T_1 = 11001,$$
$$T_2 = 10011,$$
$$T_3 = 11011,$$
$$T_4 = 00111.$$

Since T_3 has only one 0, and it is in the third position, element 3 can be eliminated, and a modified set of tests formed. This set is:

$$T_{11} = 1101,$$
$$T_{21} = 1011,$$
$$T_{41} = 0011.$$

Two additional applications of the check shows that the original set is adequate in general. Note that a set of tests may be adequate, but not adequate in general. For example,

$$T_1 = 1100,$$
$$T_2 = 0111,$$
$$T_3 = 0001$$

is a set which has this characteristic.

Probability of Failure

The introduction of information states in the testing diagram permits a more detailed specification of the status of the equipment than the simple specification of which elements are good, bad, and questionable. If sufficient data from the maintenance history of the equipment is available, the *a priori* probability that a failure is caused by a given element can be computed statistically. If such data is not available, the *a priori* probability of failure of an element can be estimated from the reliability of the component parts of the element. In the absence of either of the above, an educated guess as to the *a priori* probabilities of failure may well give a better specification of the status of the machine than the simple specification of which elements are good, bad, or questionable.

When the diagnostic problem is considered in terms of probability of failure, the influence of other factors can be introduced. The presence of one or more "symptoms" modifies, in effect, the probabilities of failure associated with the initial state. Thus, any symptoms concerning the behavior of the machine will influence the efficiency of a given testing procedure. If the symptoms indicate that element 'a' is almost certainly bad, any procedure which does not test element 'a' first will be inefficient in the presence of these symptoms. It should be emphasized that the capability of isolating bad elements by a given procedure is independent of any data on the probability of failure. However, such data can be used in the comparison of the efficiency of various diagnostic procedures.

The introduction of probability concepts also permits a generalization of the concept of a test. Previously, only tests which give unambiguous results were considered. Essentially this assumes that the testing equipment itself is not subject to failure or misinterpretation. Alternatively, the test may be considered as an operation which modifies the probabilities which specify the status of the elements. With this concept of a test, it is possible to incorporate the probability that the testing equipment is not perfectly reliable. This generalization is not considered in the remainder of this paper.

Cost

It is evident that there are a large number of possible testing procedures which can be defined for a given equipment. These procedures can only be compared if a relative rating which is indicative of the "cost" of the procedure is defined. Here cost is to be interpreted in the general sense so as to include the costs of man hours, test equipment, loss of equipment availability, etc. In the simplest model of a testing procedure, a cost can be associated with each of the possible tests. In this case, the cost of locating a particular bad element is the sum of the costs of the tests along the path which leads from the initial state to the final state indicating that this element is bad. Note that in this simple model, the cost of a given sequence of tests is independent of the order in which the tests are performed. Consideration of actual equipments indicates that very

often the cost of (*i.e.*, time required for) performing a given set of tests varies greatly as the sequence is changed. However, in the present paper, a cost is assigned to each possible test independent of what tests have been done previously.

Optimization Criteria

The most obvious criteria to apply in selection of the "best" of several diagnostic procedures is minimum average cost. In terms of the ideas previously introduced the average cost, $\bar{C}$ may be written as

$$\bar{C} = \sum_{j=1}^{N} \ell_j p_j ,$$

where p_j is the *a priori* probability of failure associated with the jth element, that is, the probability, before any diagnosis, that the failure of the equipment is caused by the failure of the jth element. ℓ_j is the cost of all of the tests which are necessary to isolate element j as bad in the diagnosis procedure under consideration, that is, the cost of determining that j is bad. If this cost is simply the sum of the costs of the tests we may write

$$\ell_j = \sum_{k=1}^{K_j} C_{kj} ,$$

where C_{kj} is the cost of test k which appears in the path from the initial state to the final state indicating that element j is bad. The minimum average cost criteria appears to be the most significant in most diagnostic procedures and is used almost exclusively in the remainder of this paper.

Another possible criteria is the so-called min-max criteria. With this criteria, the diagnostic procedure for which the maximum cost (maximized over the elements) is less than for all other procedures is selected as the best. With reference to previous notations, the best procedure in the min-max sense is obtained by computing the maximum ℓ_j for each procedure and selecting that procedure for which the $\ell_{j\ max}$ so obtained is smallest. In formal mathematical language, the best procedure in the min-max sense has a (maximum) cost of

$$C_{max} = (\ell_j^r \text{ max over } j) \text{ min over } r ,$$

where ℓ_j^r is the cost of determining element j is bad in the rth procedure.

PROBLEM COMPLEXITY AND SOLUTION OF SPECIAL CASES

In this section a detailed examination of the consequences of the definition of a test and the statement of the optimization problem is made. The work that follows is restricted by the assumption that only one element is bad. Some of the properties of the testing diagram are examined and the solution of special cases is indicated.

Problem Complexity

As a first project, it is desired to determine the number of testing diagrams that exist for various cases. First assume that there are N elements in the equipment, that exactly one element is bad, and that all of the $2^{N-1} - 1$ tests can be performed. The problem is to determine how many different testing diagrams can be constructed for this case. This number is designated as $\theta(N)$. By considering the number of diagrams possible for any $K < N$, and the number of ways in which K items can be picked from a group of N, a lower bound on $\theta(N)$ can be developed and is

$$\theta(N)_{min} = \frac{1}{2} \sum_{K=1}^{N-1} \frac{N!}{K!(N-K)!} \theta(K)\, \theta(N-K). \quad (1)$$

$\theta(N)_{min}$ is plotted in Fig. 6 for values of N up to 8. A simpler form for $\theta(N)_{min}$ can be obtained by rearranging (1) into the form

$$\frac{\theta(N)_{min}}{N!} = \frac{1}{2} \sum_{K=1}^{N-1} \frac{\theta(K)}{K!} \frac{\theta(N-K)}{(N-K)!}$$

or

$$F(N) = \frac{1}{2} \sum_{K=1}^{N-1} F(K)\, F(N-K) ,$$

where

$$\theta(N)_{min} = N!\, F(N) .$$

Fig. 6 contains a plot of F(N) vs N. This curve shows that $\log_{10} F(N)$ approaches a constant slope, indicating that F(N) behaves as $C_1(\beta)^N$ for large N. From Fig. 6, $\beta = 1.78$. Thus, $\theta(N)_{min}$ behaves as $C_1 N! (1.78)^N$ for $N > 10$.

The computation of $\theta(N)$ includes, for any N, all the distinct diagrams obtained by interchang-

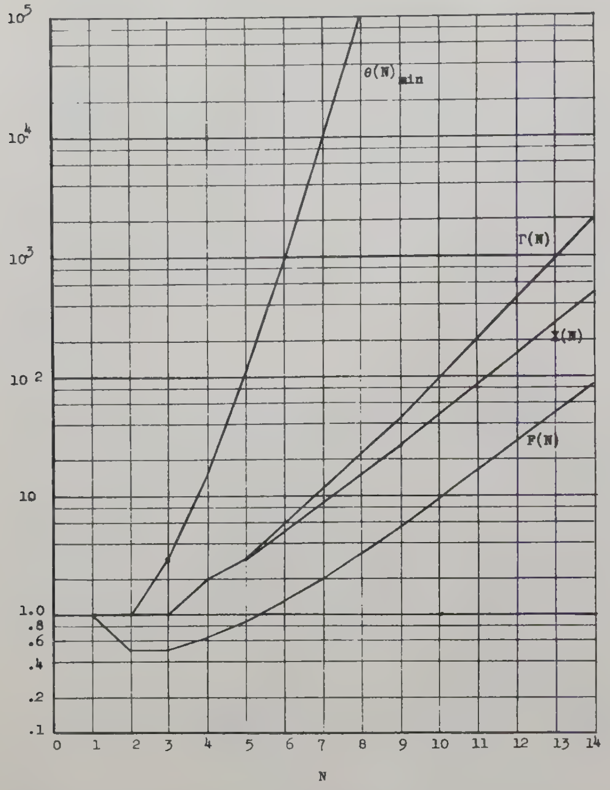


Fig. 6 —Some measures of the magnitude of the optimization problem.

ing the elements on a given structure. The number of different structures, without regard to the location of specific elements, also grows exponentially with N and is plotted as $\Gamma(N)$ in Fig. 6. In general, $\Gamma(N)$ behaves as $C_2(2.2)^N$ for large N.

In some cases examined below, attention will be focused on those structures which isolate elements after different numbers of tests. That is, it is possible for two testing diagrams to have different structures, but each will isolate 2 elements, say, in 2 tests and 4 more elements in 3 tests. The number of such different testing diagrams is denoted as X(N). Fig. 6 contains a plot of X(N) vs N. For large values of N, $\log_{10} X(N)$ is very nearly a straight line, indicating that X(N) behaves as $C_2(1.84)^N$ for $N > 10$. While X(N) is considerably smaller than $\theta(N)$ for a given N, X(N) is still quite large when it is noted that X(30) is approximately 9×10^6.

The above results lead to the obvious conclusion that it would be impractical to attempt to evaluate the testing system having the minimum average cost by the computation of the average cost associated with each. For example, with $N = 10$, there exist about 3.44×10^7 testing diagrams. Assuming that each average cost can be computed in 0.125 second, this amounts to a total time of 500 days to compute the cost associated with each of the $\theta(10)$ testing procedures. Consequently, it is apparent that it is necessary to search for some reasonable restrictions that can be imposed on this general problem, so that these restricted problems are amenable to solution by analytical means. These restrictions are the subject of discussion in the next section.

Solution of Special Cases

Equal Cost—Equal Probability: The general problem studied in the previous section can be simplified considerably if certain special cases are considered. There exists in the literature several proofs that the so-called "half-split" technique of testing results in the minimum average cost when the probability of failure is the same for all elements and the costs of all tests are the same. The half-split technique implies that at each stage in the testing procedure, a test is made that separates the remaining elements into two groups containing equal numbers of elements. It can also be shown [7] that the testing procedure is somewhat more general than this. That is, if an equipment has N elements, where

$$N = 2^m + R$$

with $0 \leqq R < 2^m$, and it is known that exactly one element is bad, then the following testing procedure yields the optimum solution:

> At each state in the sequence where there are $N' = 2^{m'} + R'$ questionable elements, choose any test which partitions the elements into two groups such that *each group* contains *at least* $2^{m'-1}$ questionable elements.

The repeated applications of this rule will yield the optimum testing procedure. Note that there are many different testing procedures that can yield the same result.[3] The average cost $\overline{C}_0$

[3] This rule yields not only the testing procedure with the minimum average cost, but also the minimum of the maximum possible cost, regardless of which element has failed. It is also the min-max solution when the probabilities of failure are unequal.

associated with the optimum testing procedure is $\overline{C}_o = (m + 2R/N)\,C_t$ where C_t is the cost of a single test.

Equal Cost—Unequal Probability: This work is now generalized to include the situation where the cost of all tests are the same, but the probability of failure of the elements may all be different. The problem that must be solved is to determine that testing procedure that yields the minimum average cost to locate the bad element. For the given N elements, and their probabilities of failure, there exist $\theta(N)$ testing diagrams. However, not all of these diagrams need to be considered in order to find the optimum. Note that if the element identification is removed from the blocks in the testing diagram, then the resulting structure is one of the $\Gamma(N)$ different structures possible. It is apparent that if two different structures require the same number of tests to isolate each of the elements, then the average cost associated with the two testing procedures will be the same. Consequently, not all of the $\Gamma(N)$ structures yield different average costs—only $X(N)$ are different.

The problem to be solved now is the determination of which of these $X(N)$ structures yield the minimum average cost. One approach would be to compute the average cost associated with each structure, and choose the smallest. However, an optimum coding problem as solved by Huffman [5] gives the solution directly. In order to show that this is true, it is necessary to draw the proper analogies between the two problems. One such set of analogies is:

Optimum Coding Problem	Optimum Testing
1) Message is sent.	1) Element is bad.
2) Message ensemble.	2) Equipment.
3) Two coding symbols.	3) Two results from a test.
4) Number of symbols in a message code.	4) Number of tests to isolate an element.
5) Time for transmission of a coded message directly proportional to the number of symbols associated with it.	5) All tests of equal cost.
6) Lowest possible average message length (optimum code).	6) Lowest possible average cost to locate the bad element (optimum testing procedure).

Using these analogies, Huffman's method to determine the optimum code yields the optimum testing procedures. The steps in the procedure are presented below, without proof, as adapted from Fano [6].

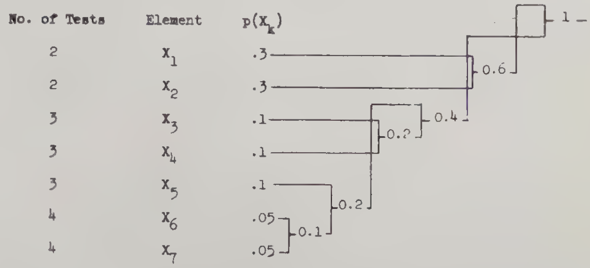


Fig. 7 —Construction of the optimum number of tests.

Step 1) Arrange the elements in order of decreasing probability, as shown in Fig. 7.

Step 2) Group together the 2 least probable elements and compute the total probability of such a subset.

Step 3) Obtain an auxiliary ensemble of elements from the original ensemble by considering the subset of 2 elements formed in Step 2 as a single element with probability equal to the probability of the subset. Rearrange this auxiliary ensemble in order of decreasing probability as shown in Fig. 7.

Step 4) Form successive auxiliary ensembles by repeating Step 2 and Step 3 until a single element of unity probability is left in the ensemble, as illustrated in Fig. 7.

Step 5) The number of tests that must be conducted to isolate each element can be determined from the tree diagram that is obtained at the conclusion of Step 4. To obtain this number, trace the path from an element to the vertex of the tree. The number of times the element is combined with other elements is the required number of tests. These results are listed in Fig. 7.

Step 6) Any testing diagram which isolates the elements in the same number of tests as determined in Step 5 will have the lowest possible average cost. One such testing diagram is shown in Fig. 8 for the example of Fig. 7. The average cost for this example is:

$$\overline{C} = C_t\,[0.3 \times 2 \times 2 + 0.1 \times 3 \times 3 + 0.05 \times 4 \times 2]$$

$$\overline{C} = 2.5\,C_t$$

This is the lowest possible average cost.

This six step procedure is the general solution

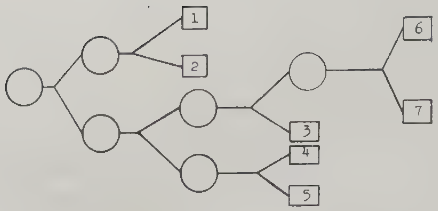


Fig. 8 —An optimum testing procedure.

for the problem where all tests have the same cost and all tests are possible. It thus includes as a special case the situation when the probabilities are all equal.

It is possible to exploit further the analogy between the optimum coding problem and the diagnosis problem. In terms of the coding problem, the following bounds exist for a binary code:

$$H(X) \leq N^* < 1 + H(X), \qquad (2)$$

where N^* is the average number of symbols per message and

$$H(X) = -\sum_{j=1}^{N} p_j \log_2 p_j \, .$$

Thus, by the analogy we have stated,

$$C_t\,[H(X)] \leq \bar{C} < C_t\,[1 = H(X)]\,. \qquad (3)$$

Eq. (3) gives upper and lower bounds on the average cost for this case. The importance of the lower bound is recognized by noting that if a testing procedure is devised under the constraint that not all tests are possible, then the average cost of this testing procedure can be compared with $C_t[H(X)]$ to determine just how much improvement would be possible if all tests were available.

CONCLUSIONS

The primary result of the present paper is the introduction of several general concepts which are applicable to the problem of diagnosis of equipment failures. Within the framework of the definitions given for an equipment, an element and the model of a test, it is possible to define and construct a diagram of a testing procedure. While the diagram has only been used here for tests which have only two possible results (pass or fail), the generalization to more detailed tests which may have several possible results is apparent. It is shown that a completely general testing diagram becomes quite complicated even when the equipment under consideration is not complicated. However, a major simplification is obtained by introducing a simplified diagram with suitably restricted tests. This simplified testing diagram may be used repeatedly in order to find all the bad elements of the equipment.

In terms of the testing diagram, it is possible to compute the minimum average cost of diagnosing the equipment, and this appears to be the most useful measure of the efficiency of a test procedure. Specific solutions for optimum procedures are included here for two special cases. However, the determination of optimum procedures for the more general case, where the costs of tests may be unequal and only a limited set of tests are available, remains unsolved. A companion paper (see footnote 2, page 26) presents a method, using information theory concepts, which leads to near optimum testing diagrams for the general problem. The combination of the solutions for special cases and the information theory method provides a useful guide for constructing efficient test procedures.

It is believed that the concepts and techniques developed here are applicable to the diagnostic procedures for a wide class of equipments; however, the ultimate evaluation can only be made after applications have been attempted for several different types of equipment. One such application is currently being attempted by the authors.

REFERENCES

[1] A. J. Hoehn and E. Saltz, "Mathematical models for determination of efficient troubleshooting routes," IRE TRANS. ON RELIABILITY AND QUALITY CONTROL, no. PGQRC-13, pp. 1-14; July, 1958.

[2] R. B. Miller, J. D. Foley, Jr., and R. P. Smith, "Systematic Trouble-shooting and the Half-Split Technique," Human Resources Res. Cen., Lackland Air Force Base, Tech. Rept. 53-21; July, 1953.

[3] L. M. Stolurow, G. Bergum, T. Hodgson, and J. Silve, "The efficient course of action in 'troubleshooting' as a joint function of probability and cost," *Educ. and Psychol. Meas.*, vol. 15, pp. 462-477; 1955.

[4] "Synchro troubleshooting," *Electronic Ind.*; June, 1958.

[5] D. A. Huffman, "A method for the construction of minimum redundancy codes," PROC. IRE, vol. 40, pp. 1098-1101; September, 1952.

[6] R. M. Fano, "Notes for Subject 6.574, Statistical Theory of Information," M.I.T., Cambridge, Mass., pp. III-10 to III-14; 1954.

[7] R. A. Johnson, E. Kletsky and J. D. Brulé, "Diagnosis of Equipment Failures," SURI, Rept. No. EE 577-594T1; April, 1959. AD-213876.

AN INFORMATION THEORY APPROACH TO DIAGNOSIS*

R. A. JOHNSON†

ABSTRACT

In the preceding paper, a model of a sequential diagnostic test procedure is developed for application to fault location in electronic equipment. The average cost of diagnosis is defined and the problem of finding procedures of minimum average cost is solved for two special cases. In the present paper, the ratio of the average information gained by performing a given test to the cost of the test is introduced as a figure of merit for the test. Repeated application of this figure of merit to choose successive tests results in a systematic way of constructing efficient sequential test procedures. The procedures so constructed are compared with known optimum procedures for several special cases as a means of evaluating the information theory approach. For some special cases the true optimum is obtained; in others, the information theory test procedures are only slightly less efficient than the optimum. The advantage of the information theory approach lies in the fact that it results in a simple systematic way of constructing efficient test procedures even in the general case for which the true optimum solution has not been obtained by other means.

*In order to maintain continuity of the preceding paper by J. D. Brule, R. A. Johnson, and E. J. Kletsky, "Diagnosis of Equipment Failures," this issue, p. 23, the above abstract is presented here. This paper, "An Information Theory Approach to Diagnosis," will be provided at the Sixth National Symposium on Reliability and Quality Control, January 11-13, 1960, and will be available in the Proceedings of that Symposium.

† Elec. Engrg. Dept., Syracuse University, Syracuse, N. Y.

RELIABILITY OF PARALLEL ELECTRONIC COMPONENTS

H. WALTER PRICE†

Summary—Electronic components are frequently connected in parallel as a measure to increase reliability. Whether the result of such a parallel connection results in an increase or a decrease in reliability, and the amount of such increase or decrease, is a function of the open-circuit failure probability and the short-circuit failure probability. Equations are derived which permit a determination of the increase or decrease of reliability when components are connected in parallel. Some curves are included to aid the circuit designer in this determination.

†Reliability Branch, Diamond Ordnance Fuze Labs., Washington 25, D. C.

$\underline{n}$ COMPONENTS CONNECTED IN PARALLEL

Electronic components may fail catastrophically by open-circuiting or by short-circuiting. A given component, then, will have a probability of failure by open-circuiting and a probability of failure by short-circuiting. The total probability of failure of such a component is given by the addition law of probability. Since, in general, a component cannot fail simultaneously by open-circuiting and by short-circuiting, the mutually exclusive event form of the law of addition of probability applies. Thus,

$$q = r + s \tag{1}$$

where q = total probability of failure
r = probability of open-circuiting
s = probability of short-circuiting

$$0 \leq r \leq 1,\ 0 \leq s \leq 1,\ 0 \leq (r + s) \leq 1.$$

Let n components be connected in parallel. These components have open-circuit probabilities of

$$r_1, r_2, r_3, \ldots, r_n$$

and short-circuit probabilities of

$$s_1, s_2, s_3, \ldots, s_n .$$

Let it be assumed that the failure probabilities are statistically independent (*i.e.*, the failure of one component in no way affects the probability of failure of the other components).

Consider, first, the case when

$$s_1 = s_2 = s_3 = \ldots = s_n = 0 .$$

Since, in this case, the components can fail only by open-circuiting, the failure of one component does not constitute a total circuit failure. In fact, all components must fail to constitute a total circuit failure. The probability of total circuit failure, then, is the joint probability of the components. Thus,

$$q_n = r_1 \cdot r_2 \cdot r_3 \cdot \ldots \cdot r_n = \prod_{i=1}^{n} r_i . \tag{2}$$

Next, consider the case when

$$r_1 = r_2 = r_3 = \ldots = r_n = 0.$$

Since, in this case, the components can fail only by short-circuiting, the failure of any one constitutes a total circuit failure. The probability of total circuit failure, then, is given by the law of addition of probability for n events. Thus,

$$q_n = 1 - \prod_{i=1}^{n} (1 - s_i) . \tag{3}$$

In the general case, where the components can fail by either open-circuiting or by short-circuiting, the probability of total circuit failure is given by the sum of (2) and (3). Thus,

$$q_n = \left[\prod_{i=1}^{n} r_i \right] + 1 - \prod_{i=1}^{n} (1 - s_i) . \tag{4}$$

If the components which connected in parallel are identical-type components, then

$$r_1 = r_2 = r_3 = \ldots = r_n$$
$$s_1 = s_2 = s_3 = \ldots = s_n ,$$

and (4) reduces to

$$q_n = r^n + 1 - (1 - s)^n \tag{5}$$

where q_n = resultant failure probability of identical-type components connected in parallel
n = number of such components parallel
r = open-circuit failure probability of each such component
s = short-circuit failure probability of each such component.

TWO COMPONENTS CONNECTED IN PARALLEL

The case of two identical-type components connected in parallel is important because of its common occurrence. For this case, (5) reduces to

$$q_2 = r^2 + 2s - s^2 . \tag{6}$$

It is of interest to know the amount of improvement in reliability by connecting the two components in parallel. An improvement in reliability is equivalent to a reduction in the probability of failure. Let there be defined, then, a failure probability reduction factor Γ_n relating the failure probability of n components in parallel to the failure probability of a single unit. Thus,

$$q_n = \Gamma_n q_1 \quad \text{or} \quad \Gamma_n = \frac{q_n}{q_1} . \tag{7 and 8}$$

It should be noted that when $\Gamma_n < 1$, the failure probability is *reduced* by the parallel connection. Conversely, when $\Gamma_n > 1$, the failure probability is *increased* by the parallel connection. Likewise, when $\Gamma_n = 1$, the failure probability is unchanged by the parallel connection. It is obvious, then, that in order to improve reliability the designer should connect the components in parallel only when $\Gamma_n < 1$.

Since q_n and q_1 of (6) are functions of r and s [see (1)], then Γ_n must be a function of r and s. In the two-component case being considered, (8) can be written

$$\Gamma_2 = \frac{q_2}{q_1} = \frac{r^2 + 2s - s^2}{r + s} . \tag{9}$$

Let a failure ratio factor η be defined as

$$\eta = \frac{\text{short-circuit probability}}{\text{open-circuit probability}} = \frac{s}{r} . \tag{10}$$

Then (9) can be written as

$$\Gamma_2 = \frac{(1 - \eta^2)r + 2\eta}{1 + \eta} \tag{11}$$

such that Γ_2 is a function of the open-circuit probability r and the ratio of the short-circuit to open-circuit probabilities η.

Fig. 1 is a chart for use by a circuit designer to determine the change in probability of failure by connecting two identical-type components in parallel. This chart was computed from (11) and gives Γ_2 as a function of r with η as a parameter.

In Fig. 1 it is to be noted that

$$\Gamma_2 = 1 \quad \text{when} \quad \eta = 1$$

for all values of r. Thus, if the short-circuit probability is equal to the open-circuit probability, the total failure probability of the two components in parallel is identical to the failure probability of a single unit. Hence, the curve for $\eta = 1$ has been labeled the "break-even" line in Fig. 1.

From Fig. 1 and (11) it is obvious that

$$\left.\begin{array}{l} \Gamma_2 < 1 \text{ when } \eta < 1 \\ \Gamma_2 = 1 \text{ when } \eta = 1 \\ \Gamma_2 > 1 \text{ when } \eta > 1 \end{array}\right\} \begin{array}{l} \text{for all values of} \\ r \text{ [subject to the} \\ \text{restrictions} \\ 0 \leq r \leq 1, \\ 0 \leq (r + s) \leq 1]. \end{array}$$

Therefore, to determine if a reliability improvement can be realized by connecting two components in parallel, it is only necessary to determine whether

$$\eta < 1 \quad \text{or} \quad \eta \geq 1.$$

The magnitude of such an improvement (or degradation) can then be determined with the aid of Fig. 1.

THREE COMPONENTS IN PARALLEL

For the case of three identical-type components, (5) reduces to

$$q_3 = r^3 = 3s - 3s^2 + s^3 \tag{12}$$

and (8) reduces to

$$\Gamma_3 = \frac{q_3}{q_1} = \frac{r^3 + 3s - 3s^2 + s^3}{r + s}. \tag{13}$$

Eq. (13) can be written in terms of η from (8) as

$$\Gamma_3 = \frac{(1 + \eta^3)r^2 - 3\eta^2 r + 3\eta}{1 + \eta}. \tag{14}$$

Fig. 2 is a chart for use by a circuit designer to determine the change in probability of failure by connecting three identical-type components in parallel. This chart was computed from (5) and gives Γ_3 as a function of r with η as a parameter.

In Fig. 2 it can be seen that

$$\Gamma_3 \simeq 1 \quad \text{when} \quad \eta = \tfrac{1}{2}$$

for most values of r. Therefore, the curve $\eta = \frac{1}{2}$ has been labeled the "break-even" line in Fig. 2.

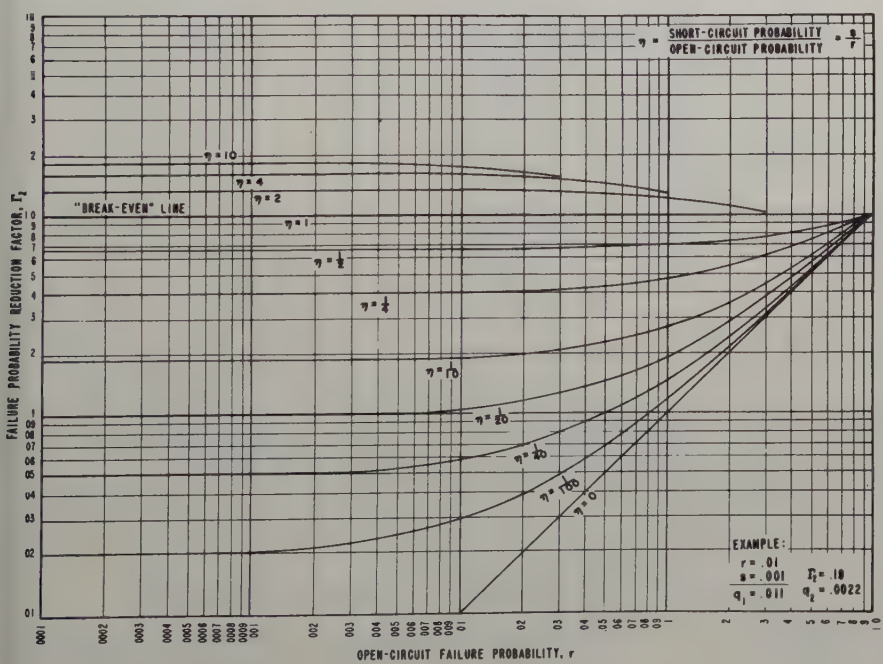


Fig. 1—Failure probability reduction factor (Γ_2) for 2 identical-type components in parallel.

It is obvious, then, that for three components in parallel

$$\left.\begin{array}{l} \Gamma_3 < 1 \quad \text{when} \quad \eta < \frac{1}{2} \\ \Gamma_3 = 1 \quad \text{when} \quad \eta = \frac{1}{2} \\ \Gamma_3 > 1 \quad \text{when} \quad \eta > \frac{1}{2} \end{array}\right\} \begin{array}{l} \text{for most values of} \\ \text{r [subject to the} \\ \text{restrictions} \\ 0 \le r \le 1, \\ 0 \le (r + s) \le 1]. \end{array}$$

As in the two component case, it is only necessary to examine the value of η to determine if a reliability improvement can be made by connecting three components in parallel.

OPTIMUM NUMBER OF PARALLEL COMPONENTS

In Figs. 1 and 2, it can be seen that 3 components in parallel results in a lower value of Γ than 2 components for certain combinations of values of r and η. However, 2 components in parallel results in a lower value of Γ than 3 components over a much larger combination of values of r and η. It follows then that for any particular combinations of values of r and η there is an optimum number of components connected in parallel which will result in the lowest value of Γ and, hence, result in the highest reliability circuit.

Consider, then, that there are regions related to certain combinations of values of r and η where 2 components in parallel constitute the most reliable circuit. Similarly, there are regions where 3 components in parallel constitute the most reliable circuit, etc.

The regions corresponding to the different numbers of optimum components are adjacent to each other such that the region corresponding to 2 components lies between the regions corresponding to 1 component and 3 components, etc. Two adjacent regions will be separated by a border line.

To obtain the location of these border lines, consider the general form for Γ_n.

$$\Gamma_n = \frac{r^n + 1 - (1 - s)^n}{r + s} . \tag{15}$$

To determine the border line between the 1 component and 2 component regions set

$$\Gamma_1 = \Gamma_2 .$$

Solving for η, it can be determined that $\Gamma_1 = \Gamma_2$ when $\eta = 1$ for all values of r.

Likewise, to determine the border line between the 2 component and 3 component regions set

$$\Gamma_2 = \Gamma_3 .$$

This results in an equation of η and r which

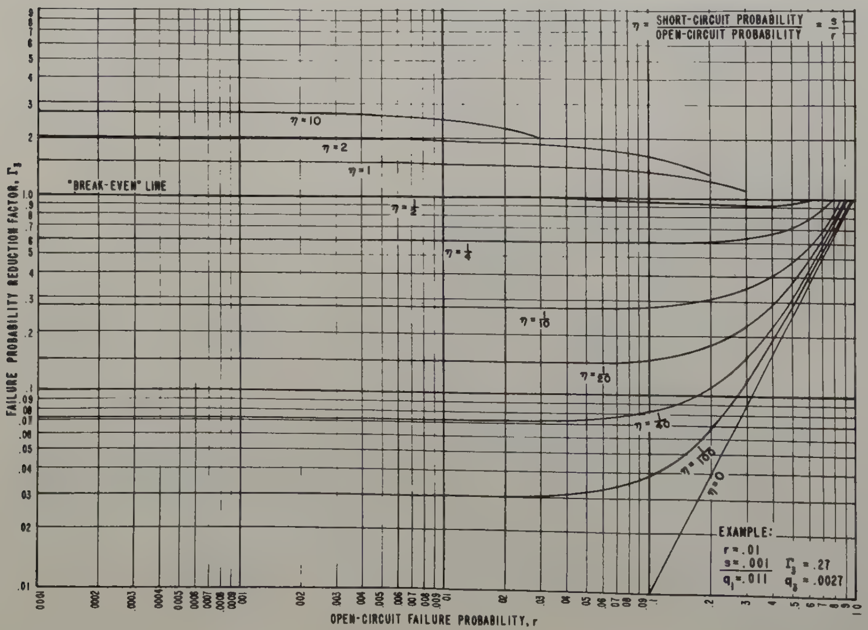


Fig. 2—Failure probability reduction factor (Γ_3) for 3 identical-type components in parallel.

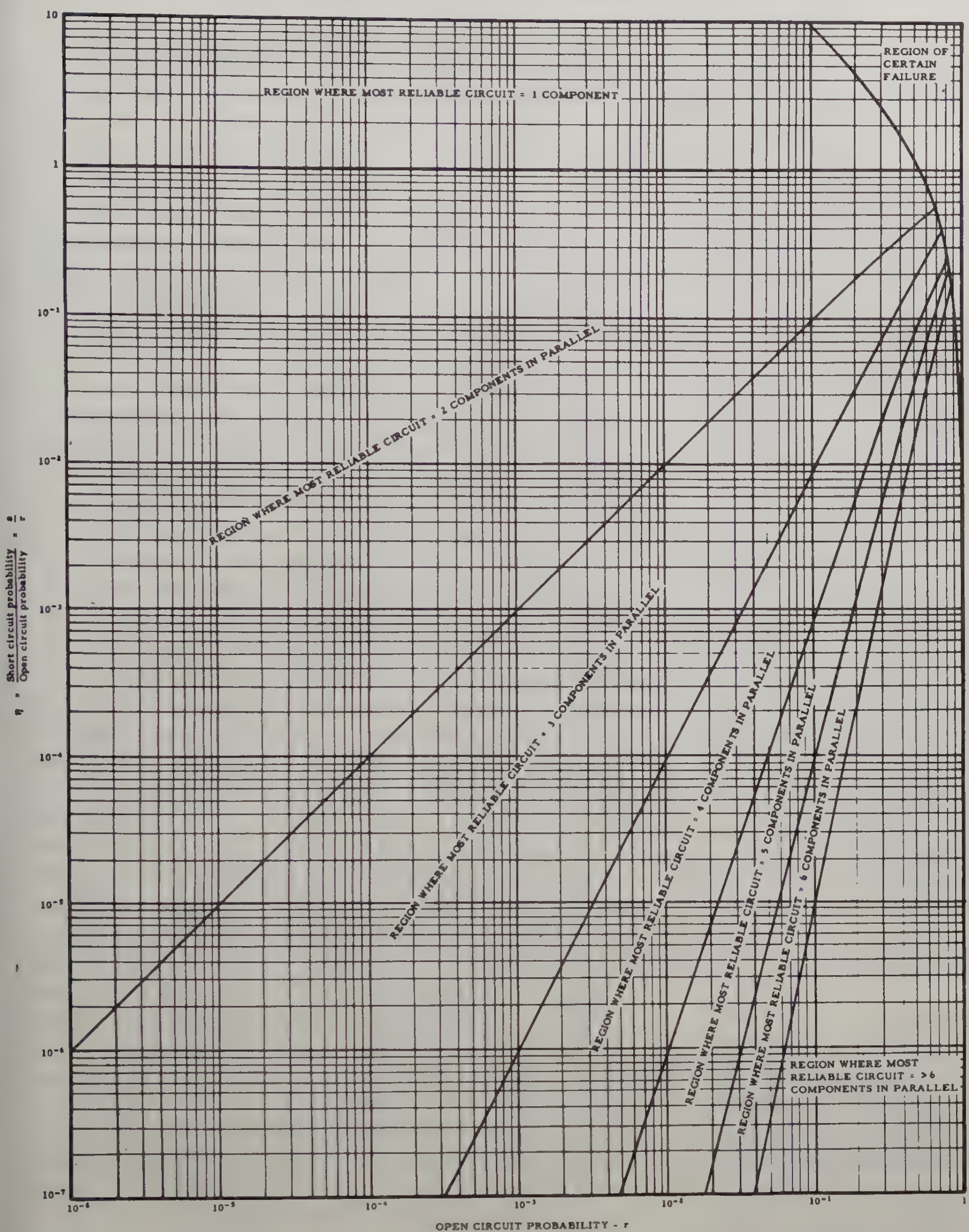


Fig. 3—Optimum number of parallel components.

is the equation of the border line separating the regions.

Fig. 3 shows the results of computing the border lines for each of the regions up to 6 components in parallel. It is to be noted that 1 component is the most reliable for all $\eta > 1$. It is to be further noted that 2 components in parallel results in the most reliable circuit for most practical values of η and r for $\eta < 1$.

CIRCUIT DESIGN PROCEDURE

1) Obtain values (or estimates) for s and r.
2) Use Fig. 3 to determine the most reliable number of components to place in parallel.
3) Use Figs. 1 or 2 or

$$\Gamma_n = \frac{r^n + 1 - (1 - s)^n}{r + s}$$

as may be applicable to determine the reduction in failure probability.

4) Use

$$q_n = \Gamma_n q_1 = \Gamma_n(r + s) = \Gamma_n(r + \eta r)$$

to determine the total failure probability of the parallel components.

EVALUATION AND PREDICTION OF CIRCUIT PERFORMANCE BY STATISTICAL TECHNIQUES

J. MARINI and R. WILLIAMS†*

Summary—A method is described for predicting circuit performance to the extent that it is dependent on part performance. The basis for the prediction is the performance of parts as measured at test points fixed by the part specifications. Implicit in this method is the assumption that the distribution of part performance at the test points can be predicted from consideration of the specifications. Such an assumption is necessary to any attempted prediction of this nature.

An empirical equation giving circuit performance in terms of part performance as measured at the test points is assumed. The exact form of the equation is determined experimentally, by means of regression analysis of data consisting of sets of measurements of breadboard models of the circuit. The empirical equation is then used mathematically to calculate the distribution of the circuit performance from the assumed distributions of part performance.

The method has been applied successfully to predict the laboratory performance of an ac amplifier and a telemetering oscillator. In principle, the method can be extended to the prediction of equipment or system performance.

I. INTRODUCTION

The evaluation of the design of a circuit intended for mass production can be regarded as a problem in prediction. The conclusion that a given circuit design will be satisfactory actually amounts to a prediction that the performance of most of the circuits to be produced under the design will fall within specified limits. In this sense, the preproduction evaluation of a circuit design can be regarded as a statistical prediction of the distribution of circuit performance.

In practice, this prediction usually is based to a considerable extent on observations of the performance of a number of preproduction models of the circuit. Occasionally, the performance may be observed when "limit" tubes or parts (that is, tubes or parts whose characteristics barely meet specification limits) are used. However, merely to observe the performance of the models without simultaneously relating this performance to the properties of the particular parts which are used to construct the models can be misleading. The use of limit tubes and parts is a crude attempt to avoid this danger.

The purpose of this paper is to describe a systematic, quantitative method of predicting the initial performance of an electronic device on the basis of the initial performance of its parts, through use of the statistical technique of multiple-regression analysis. An important feature of this method is that it provides a means of bridging the gap between circuit performance and specification-controlled part performance.

The method is demonstrated by means of the following problem: an oscillator which is to be produced in quantity has been designed to operate at a certain frequency. It is known that, when a number of these oscillators are constructed, the output frequencies will be distributed over a range of values. The variations from the desired value will be due principally to variations in the performance characteristics of the parts used in making the oscillators. If the design is good, the mean of the distribution will fall near the design-center frequency. However, unless it is practical to adjust the frequencies of individual oscillators, it may be necessary to reject those in which the deviations from design center are too wide. Therefore, it would be highly desirable to be able to predict what the probability distribution of the output will be when the oscillators are manufactured in large numbers. This would make it possible to determine in advance what percentage of the oscillators would be acceptable.

Formulation of the Problem

In the method reported here, it is assumed that the performance of a circuit can be expressed in terms of one or more measurable variables, which are termed "circuit characteristics"—for example, the frequency and power output of an oscillator. In order to simplify this exposition, only one characteristic of the oscillator circuit—frequency—will be considered. However, the methods used can be extended to handle more than one characteristic.

It is also assumed that the performance of parts can be expressed in terms of measurable

*Electromagnetic Res. Corp., Washington, D. C.; formerly at Arinc Res. Corp., Washington, D. C.
†Arinc Res. Corp., Washington, D. C.

variables called part characteristics—*e.g.*, the resistance of a resistor and the transconductance of an electron tube. In this connection, however, it is important to observe that the value obtained for a given part characteristic depends on the conditions under which it is measured. The term "operating part characteristic" will be used here to designate a characteristic measured under actual circuit operating conditions, while the term "specification part characteristic" will be used to signify a characteristic measured under the conditions specified in the part-specification sheet. Unless the conditions under which a characteristic is measured in the circuit are identical with those stipulated in the specification, the operating part value and the specification part value of a measured characteristic are most probably different.

The fact that operating part characteristics and specification part characteristics are not identical gives rise to some difficulties. Design handbooks ordinarily give equations for circuit characteristics in terms of operating part characteristics.[1] Therefore, if a designer wishes to predict the distribution of the circuit-characteristic values which will result from his design, he must have information concerning the distribution of the *operating* characteristics of the parts used in his design. At best, however, only nominal or design-center values of the operating part characteristics will be provided by the manufacturer. The only good source of information about the distribution of part characteristics is the part specification; but, to make use of these specifications, it is necessary to bridge the gap between the circuit characteristic and the *specification* part characteristic.

Stated in mathematical terms, then, the problem under consideration is prediction of the probability distribution, $p(Y)$, of a circuit characteristic, Y, where Y is considered to be a function of the *specification* part characteristics.

[1] For example, the gain at resonance of a loaded tuned amplifier may be given as a function of the transconductance of the tube and the resistance of the loading resistor. However, the value of transconductance used in the equation is measured at the operating point of the tube as it is used in the circuit, and this, in general, is different from the value that would be obtained through measurement at the specification test point. Similarly, the resistance value given in the equation is the equivalent shunt resistance of the resistor as measured at the operating frequency of the circuit. If the frequency is high, this value will differ from the dc resistance prescribed in the specifications.

Actually, the method described in this paper is used to predict, not $p(Y)$, but only the mean, μ, and the variance, σ^2, of the distribution of Y. Fortunately, this is not a very severe limitation, because, where the distribution of Y is normal, the exact probability distribution is determined by these quantities, and where it is not normal, a great deal of information is provided by them. It is known that 95 per cent of the population of Y must lie within the limits $\mu \pm 2\sigma$ when Y is normally distributed, and it can be shown that 95 per cent of the population of Y must lie within the limits $\mu \pm 4.5\sigma$, no matter what the shape of the distribution happens to be.[2]

Illustrative example: The method used in this paper has much in common with a technique described in a Wright Air Development Center Report[3] for predicting tolerance limits on the output of a circuit whose performance depends on only a single specification part characteristic. The example given in the report is ideal for illustrating the basic ideas underlying the method used here, and is reproduced below. The circuit characteristic of interest was the output current of an electron tube as measured in the circuit, and the only specification part characteristic of importance was the plate current of the tube at a test point fixed by the specifications. When a number of tubes were measured at the specification test point, and then inserted consecutively into the circuit, the data shown in Table I were obtained. These data are plotted in Fig. 1. The straight line drawn in Fig. 1 is determined by the least-squares solution of the equation

$$I_C = b_0 + b_1 I_S + \epsilon$$

where I_C is the tube output current as measured in the circuit, and I_S is the current measured at the specification test point. The vertical scatter of the points about this line corresponds to the error term, ϵ.

Fig. 1 suggests a number of considerations involved in the use of this type of method. First, the distribution of test-point plate currents of the sample of tubes used does not represent very well the distribution allowed under the specification. Second, the illustration involves extrapolation, although this would not have been necessary if

[2] A. Hald, "Statistical Theory with Engineering Applications," John Wiley and Sons, Inc., New York, N. Y., pp. 109-110; 1952. (Tchebycheff's Inequality.)

[3] Author, "Techniques for Application of Electron Tubes in Military Equipment," Wright Air Dev. Center, Dayton, Ohio, Tech. Rept. 55-1; October, 1955.

TABLE I

EXPERIMENTAL DATA ON ELECTRON-TUBE OUTPUT

Tube Number	Plate Current (milliamperes)	Circuit Current (milliamperes)
1	8.10	4.54
2	8.90	4.90
3	7.40	4.29
4	7.00	4.14
5	7.80	4.54
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
50	7.60	4.44

tubes were selected to cover the entire permissible range of variation of test-point current. Third, the fact that the scatter of the points about the least-squares line is small makes it seem likely that the output current is indeed determined by the specifications.

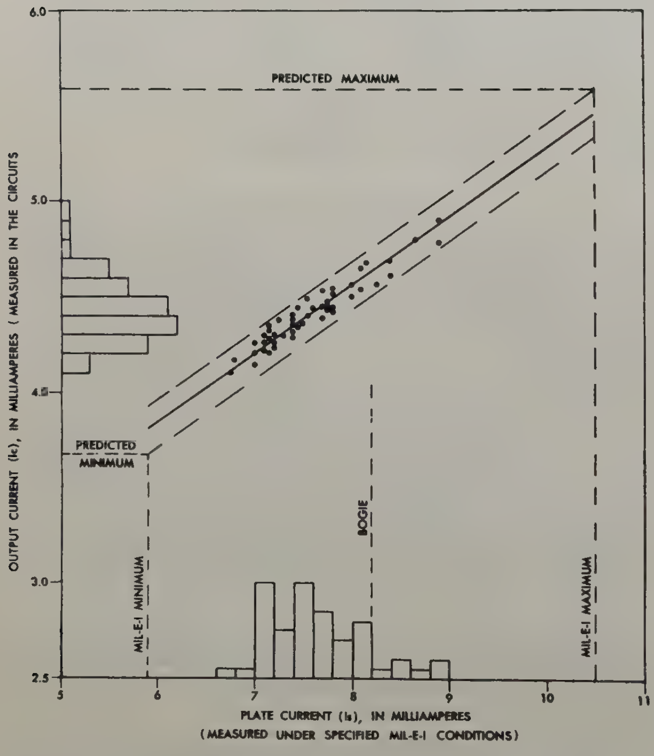


Fig. 1—Circuit output vs specification current, showing regression line used for prediction.

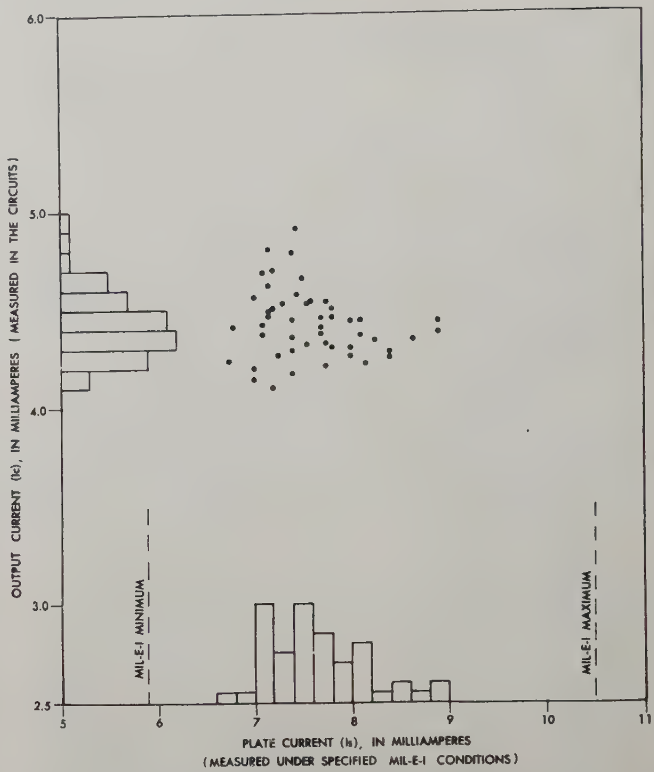


Fig. 2—Circuit output vs specification current—not suitable for prediction.

Fig. 2 was drawn to illustrate the fallacy of merely measuring the circuit characteristic, without analyzing the relationship between it and the specification part characteristics. The results shown in Fig. 2 might conceivably have been obtained on the same circuit that provided the data shown in Fig. 1. In Fig. 2, it is evident that the output is not controlled by the specification part characteristic selected, and that further study is required to determine the source of the variation of the circuit characteristic. If the circuit characteristic is dependent on part characteristics not controlled by the specifications, there is always the possibility of large shifts in the distribution of the circuit characteristic when different lots of parts are used.

This possibility exists, although to a lesser degree, even when good correlation is obtained. For example, a circuit may be critical with respect to grid current, while the tubes used in the experiment may happen to be from a lot which is exceptionally good in this respect. It should be borne in mind that the predictions made here are merely for the best possible behavior to be expected under the specifications over a long period of time. Extraneous factors not included in the considerations can always arise to invalidate the

predictions. However, a quantitative prediction of performance as a guide makes possible the detection of these extraneous factors when they do occur.

In general, the method described in this paper is a mathematical extension of the basic ideas illustrated in Fig. 1. However, when more than one part characteristic is considered, it is no longer satisfactory to use the concept of minimum and maximum limits. Instead, it becomes necessary to resort to the distribution concept.[4]

II. METHOD

Description

This paper describes a method for predicting the values of the mean, μ, and the variance, σ^2, of the distribution of a circuit characteristic, Y, by using regression analysis to determine empirically a relationship between Y and the specification part characteristics, $X_1 \ldots X_k$, of the parts used in the circuit. The determining relationship and the distribution of the specification part characteristics as assumed from knowledge of the specification are used to calculate predicted values for μ and σ^2. The general method is to assume that the equation which relates the circuit characteristics to the specification part characteristics can be expressed as a linear combination of known functions, $f_i(X_1, X_2, \ldots X_k)$, of the specification part characteristics[5] $(X_1, X_2, \ldots X_k)$, plus a random variable, ϵ.

$$Y = b_o + b_1 f_1(X_1, X_2 \ldots X_k) + \ldots + b_h f_h(X_1, X_2 \ldots X_k) + \epsilon. \tag{1}$$

It is assumed that the functions f_i can be so chosen that practically all of the variation in Y attributable to the specification part characteristics $X_1 \ldots X_k$ will be contained in the linear combination. The random variable ϵ will then represent the variation in Y caused by variation in operating part characteristics, experimental error, and other sources not controlled by the values of the specification part characteristics considered. It is arbitrarily assumed that ϵ is normally and independently distributed, with mean zero and an unknown variance, σ_ϵ^2.

If a number of models of the circuit are constructed and the values of $X_1 \ldots X_k$, together with the corresponding value of Y, are measured on each model, it is possible to solve (1) for the values of the unknown constants $b_o \ldots b_h$, and for the unknown variance σ_ϵ^2, by means of regression analysis. The values of the constants and the variance can then be used to estimate the values of μ and σ^2. Expressions for μ and σ^2 in terms of $b_o \ldots b_h$ and σ_ϵ^2 can be obtained from (1) by using the properties of the expected value[6] and following the procedures described below. An illustration or two should clarify this.

Linear expansion: First, assume that Y can be expanded in terms of the specification part characteristics in a Taylors' series with second-order and higher terms neglected. Eq. (1) would then become

$$Y = b_o + b_1 X_1 + \ldots + b_k X_k + \epsilon.$$

Taking the expected value of Y,

$$E(Y) = b_o + b_1 E(X_1) + \ldots + b_k E(X_k) + E(\epsilon) = b_o + b_1\mu_1 + \ldots + b_k\mu_k, \tag{2}$$

where μ_i is the mean of X_i. The expected or mean value of ϵ is assumed to be zero. Also,

$$\begin{aligned}\sigma^2 &\equiv E(Y - \mu)^2 \\ &= E[b_1(X_1 - \mu_1) + \ldots + b_k(X_k - \mu_k) + \epsilon]^2 \\ &= b_1^2 E(X_1 - \mu_1)^2 + \ldots + b_k^2 E(X_k - \mu_k)^2 \\ &\quad + E(\epsilon^2) + 2b_1b_2 E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &\quad + \ldots + 2b_1 E[(X_1 - \mu_1)(\epsilon)] + \ldots\end{aligned}$$

$$\sigma^2 = b_1^2\sigma_1^2 + \ldots + b_k^2\sigma_k^2 + \sigma_\epsilon^2 + 2b_1b_2\sigma_{1,2} + \ldots \tag{3}$$

where σ_i^2 is the variance of X_i, σ_ϵ^2 is the variance of ϵ, and $\sigma_{i,j}$ is the covariance between X_i and X_j. Since it has been assumed that ϵ is distributed independently of X_i, the covariance between ϵ and X_i is zero.

[4] R. C. Miles, "Tolerance considerations in electronic product design," *Electronic Design*, vol. 1, pp. 6-7, May, 1953; vol. 1, pp. 6-7, June, 1953.

[5] Other variables—*e.g.*, part characteristics not controlled by the specifications, applied voltages, and even ambient temperature—can also be included as arguments of the functions. However, the performance prediction will be improved by inclusion of such variables only if information is available concerning their probability distributions.

[6] A discussion of expected values can be found in most books on mathematical statistics. The value of a function $f(X_1 \ldots X_n)$ is defined as an integral,

$$E(f) = \int \ldots \int f\, p(X_1, X_2, \ldots, X_n)\, dX_1 dX_2 \ldots dX_n.$$

From this definition, it is easy to show that the expected value of a sum equals the sum of the expected values; that the expected value of the product of a constant and a variable equals the product of the constant and the expected value of the variable; and that the expected value of the product of independent variables equals the product of the expected values.

As was stated above, numerical values for the b's can be obtained experimentally by use of regression analysis, as can an estimate for σ_ϵ^2. Numerical values for μ_i and σ_i^2 can be obtained from the specifications.[7] The numerical value for the covariance terms $\sigma_{i,j}$ will probably have to be estimated on the basis of measurements on the parts available.[8] By substituting these numerical values into (2) and (3), the numerical values sought for the mean and variance of the circuit characteristic Y are obtained.

Nonlinear expansion: To illustrate the possibility of using nonlinear terms, assume that Y is dependent on only two specification part characteristics, and can be written, say, as

$$Y = b_o + b_1X_1 + b_2X_2^2 + \epsilon .$$

Then,

$$\mu = E(Y) = b_o + b_1E(X_1) + b_2E(X_2^2) + E(\epsilon)$$

$$\mu = b_o + b_1\mu_1 + b_2(\mu_2^2 + \sigma_2^2) .$$

Also,

$$\sigma^2 = E[(Y - \mu)^2] = E(Y^2) - 2\mu E(Y) + \mu^2 = E(Y^2) - \mu^2$$

$$\sigma^2 = E(b_o + b_1X_1 + b_2X_2^2 + \epsilon)^2 - \mu^2 = b_o^2 + b_1^2\,E(X_1^2) + b_2^2\,E(X_2^4) + E(\epsilon^2) + 2b_ob_1E(X_1) + 2b_ob_2E(X_2^2) + 2b_1b_2E(X_1X_2^2) - \mu^2$$

$$\sigma^2 = b_o^2 + b_1^2\,(\mu_1^2 + \sigma_1^2) + b_2^2E(X_2^4) + \sigma_\epsilon^2 + 2b_ob_1\mu_1 + 2b_ob_2(\mu_2^2 + \sigma_2^2) + 2b_1b_2\,E(X_1X_2^2) - \mu^2 .$$

In this case, numerical values for the fourth moment of X_2 and for the product moment[9] $E(X_1X_2^2)$ may have to be estimated from measurements of parts on hand.

From this example, it should be clear that the method used is by no means limited to an expansion involving first powers of the specification part characteristics. The only penalty for using higher powers or products of powers is the occurrence of higher-order moments in the expressions for μ and σ^2, and the consequent necessity for estimating values for these moments from the specifications—or, as a last resort, from measurements on the parts used.

III. APPLICATION OF THE METHOD

The method described in the preceding section has been applied to a voltage-controlled oscillator. A circuit diagram of the oscillator is shown in Fig. 3.

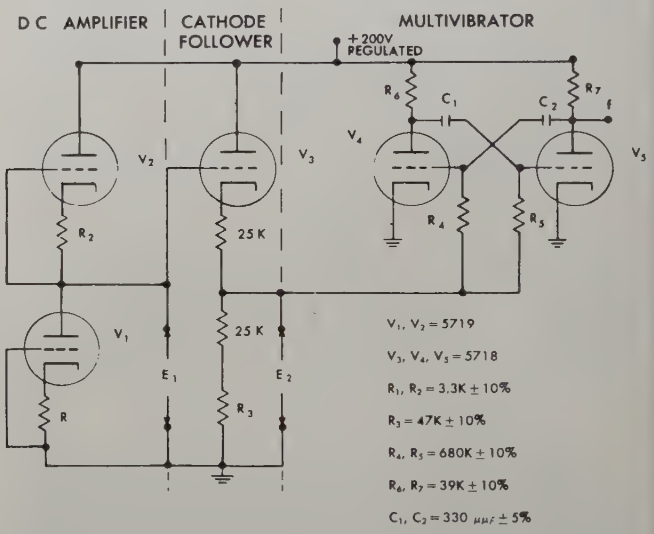


Fig. 3—Voltage controlled oscillator (grounded input).

Definition of the Problem

To apply the multiple-regression technique, it is first necessary to specify the circuit and define the performance characteristic of interest. The circuit was chosen to be that shown in Fig. 3, under the assumption of zero input and stable

[7] Often, of course, the specifications do not contain information about μ_i and σ_i, but merely give tolerance limits on the value of the specification part characteristic. Even in these cases, however, the realistic approach would seem to be to estimate these values, using the limits as a guide.[4]

[8] $\sigma_{i,j}$ can be estimated from the sample by using the formula

$$\sigma_{i,j} = \frac{\Sigma(X_i - \overline{X}_i)(X_j - \overline{X}_j)}{[\Sigma(X_i - \overline{X}_i)^2 \cdot \Sigma(X_j - \overline{X}_j)^2]^{\frac{1}{2}}}\,\sigma_i\sigma_j .$$

This formula was derived by assuming that the correlation coefficient between characteristics of the parts available in the laboratory is the same as that between characteristics of the parts in the population.

[9] These moments can be expressed in terms of moments about the mean by using the identity

$$X^n \equiv (X-\mu)^n + n(X-\mu)^{n-1}\mu + \frac{n(n-1)}{2!}(X-\mu)^{n-2}\mu^2 + \ldots + \mu^n$$

and taking expected values. In the case of a normal distribution, the fourth moment about zero reduces to

$$E(X^4) = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4$$

and this expression can be used to estimate the fourth moment from the specification.

power supply. To provide a simple illustration of the method, the circuit performance characteristic selected was the nominal output frequency under the conditions given above. The problem was to predict the distribution of this output frequency for the population (production run) of oscillators.

An important practical consideration in the application of the method to the oscillator circuit is the number of sets of observations necessary. A rough rule of thumb requires that the number of sets of observations should exceed the number of variables used in the regression equation by 30 or more. Increasing the number of observations increases the accuracy with which the regression coefficients are determined. If the rule of thumb is used, it follows that each additional part characteristic or variable included in the regression equation necessitates at least one additional set of observations. In turn, each new set of observations necessitates additional measurements on all of the part characteristics used in the regression equation. In other words, the total number of measurements required tends to increase rapidly with the number of variables used in the regression equation. For this reason, the variables to be used should be carefully selected; otherwise, the amount of laboratory work involved may become prohibitive. Since the computational work also increases rapidly with the number of variables used, an automatic computer is practically a necessity if there are more than four or five variables. However, if only two or three variables are used, this work can easily be done on a desk calculator.

Circuit Analysis

The circuit must be analyzed to determine what parts and what performance characteristics of these parts influence the circuit performance characteristic of interest. In addition to the circuit elements, external factors, such as supply-voltage variation, can also be considered. The analysis has two purposes: 1) to determine the form of the regression equation, and 2) to reduce the number of variables appearing in it. The number of variables can be reduced either by eliminating those factors which obviously have no bearing on the circuit characteristic, or by dividing the circuit into smaller sub-circuits, or stages, which can be considered individually. The latter device accomplishes the objective of reducing the number of variables in the regression equations without sacrificing a variable which might influence the circuit characteristic. As experience with this technique shows, the solution of two or more small regression equations involves less effort than the solution of one large equation with many variables. Another advantage of the breakdown of the circuit into its individual stages is that this facilitates determination of the variables that influence circuit performance. In many circuits, the stages follow one another in a series-type arrangement. When this is the situation, the variables in any given stage can be combined into one variable (the output of the stage) for inclusion in the next succeeding stage.

The oscillator circuit shown in Fig. 3 could have been treated as an entity, with simultaneous consideration given to all of the parts in the three stages. However, since this would have involved a regression equation with a large number of variables, it was considered more practical to analyze the circuit stage-by-stage. Circuit analysis resulted in dividing the oscillator unit into three stages—dc amplifier, cathode follower, and multivibrator—as indicated by the dotted lines. The dc amplifier may be considered as an independent circuit with an output performance characteristic, E_1. When the cathode follower is considered independently, E_1 must be considered a variable influencing the cathode-follower output, E_2. Similarly, when the multivibrator is treated, E_2 will be included as a variable influencing the multivibrator output frequency, f, which is, in effect, the output frequency of the oscillator. Through this procedure, all of the variables influencing the oscillator output frequency will be considered in the multivibrator regression equation, and the three regression equations that must be solved will all be much simpler than the combined equation would have been.

Even if this procedure is followed, it is still desirable, if at all possible, to further reduce the number of variables. In the case of the dc amplifier (which will be used hereafter for illustrative purposes), it is obvious that, under no-signal conditions, variability in E_1 would be due to the resistive unbalance on either side of the E_1 pick-off point. Since a stable voltage supply has been assumed, the circuit parts contributing to the unbalance would be the load resistors and the type 5719 tubes. The specification part characteristics which must be related to E_1 are believed to be the dc resistance of the load resistor and the specification plate current (I_b) of the type 5719 tubes. Since it is the unbalance, or difference, between the two resistors and tubes in the circuit that is of importance, it is logical to utilize the values of unbalance in determining the relationship. Consequently, the form of the

regression equation chosen for the dc amplifier was

$$E_1 = b_0 + b_1(I_{b_1} - I_{b_2}) + b_2(R_1 - R_2) + \epsilon \quad (4)$$

where ϵ is the error term added to take care of experimental and other variations in the data.

Experimental Solution

The problem now is to obtain data to use in solving the equation and verifying its validity. This is accomplished through a designed laboratory experiment.

Considerations involved in the design of an experiment are, of course, the size of the experiment, the accuracy of the test equipment and the methods used to obtain the data. The size of the experiment depends upon the degree of precision desired in the predicted results of the analysis and, consequently, in the estimate of the equation. Unfortunately, the precision in this case depends to a large extent on the value of the variance of the error term—and this value is not determined until the experiment has been performed and the data have been analyzed. In the experiment described here, 31 complete sets of observations were obtained and used in the analysis.

The accuracy of the test equipment will appreciably affect the precision with which the estimate of the equation can be made. The important consideration is the need to avoid a consistent bias in the test equipment. It is better to have equipment of relatively low accuracy, if the variations about the true value are randomly distributed, than to have equipment which provides better accuracy, but is consistently biased high or low. Consistent bias will show up in the equation and throw the prediction or the estimated equation off by a proportional amount.

To obtain the required data, it is necessary to construct a number of breadboards—*i.e.*, the circuit-output performance characteristics of interest must be measured in what is essentially a different circuit each time. Obviously, the economical way to do this is to make a single breadboard mockup of the circuit, and change the circuit elements under consideration each time. This device can easily be used in the case of low-frequency applications; but, when higher frequencies are involved, proper consideration must be given to the design of the breadboard or it may be necessary to construct a large number of breadboards.

Whenever possible, it is desirable to select parts that are representative of the parts that would be used in production. For example, if the tubes in the circuit are of types produced by several manufacturers, it is desirable to have a sample of tubes from each manufacturer who might provide tubes for the production run. Similarly, the parts used in a breadboard should be chosen at random from the total supply of parts available, so that they will be reasonably representative of the whole supply of parts available.

In the study of the oscillator circuit, one breadboard model was constructed with spring clamps used to hold the resistors and capacitors in the breadboard, and standard tube sockets with subminiature adaptors used for the tubes. The parts and tubes chosen for the experiment were selected at random from those available within the laboratory, and each part was numbered individually. The parts were then tested for the characteristics considered important on the basis of their applicable procurement specifications. In the case of the dc amplifier, the type 5719 tubes were measured as specified in MIL-E-1. Sets of parts were then inserted into the complete breadboard (which used regulated power supplies to insure constant supply voltages), and the output performance characteristic of interest (output frequency) was recorded. It should be noted that this experiment was performed on the complete oscillator, so that a particular set of parts throughout each stage could be related to the observations in any other stage. The oscillator frequency, f, and the outputs, E_1 and E_2, respectively, of the first two stages were recorded for each set of parts. A sample of the data obtained on the dc amplifier during the experimental portion of the method is given in Table II.

TABLE II

EXPERIMENTAL DATA ON DC AMPLIFIER

Set No.	$I_{b_1} - I_{b_2}$	$R_1 - R_2$	E_1
1	+ 0.07	+ 146	98.8
2	+ 0.03	+ 96	99.2
3	- 0.10	- 63	103.4
4	0.00	+ 62	102.8
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
31	- 0.09	+ 38	107.4

$I_{b_1} - I_{b_2}$ = difference in plate current, in milliamperes.
$R_1 - R_2$ = difference in resistances, in ohms.
E_1 = output, in volts.

Results of Regression Analysis

Regression analysis[10] of the experimental data on the dc amplifier portion of the oscillator effected a solution of (4) for the partial regression coefficients which yielded

$$\begin{aligned} E_1 &= b_0 + b_1(I_{b_1} - I_{b_2}) + b_2(R_1 - R_2) + \epsilon \\ &= 100.17 + (-56.04)(I_{b_1} - I_{b_2}) \\ &\quad + (.016)(R_1 - R_2) + \epsilon \end{aligned} \tag{5}$$

where current was expressed in milliamperes and resistance in ohms. The estimate of the variance of the error term was $\sigma_\epsilon^2 = 2.055$ volts squared.

Before use was made of (5), it was tested for significance—that is, to determine whether or not the equation adequately represented the true condition, and whether or not the individual terms in the equation contributed significantly toward explaining the variability of E_1. The test of significance for the total regression equation was set up as an analysis of variance.[11] The equation proved highly significant. However, just because the regression equation satisfactorily explained the variability of E_1, it would not necessarily follow that each term in the equation contributed significantly. It was therefore necessary to make tests of significance on the individual b values. This was conveniently accomplished by means of the "t" test, which indicated that both b_1 and b_2 were highly significant. (In this instance, the t values were $t_1 = 15.6$ and $t_2 = 6.8$, both of which are much greater than the critical value, which is $t = 2.048$ for 28° of freedom and a 0.05 level of significance.) Consequently, it can be accepted that (5) represents a suitable mathematical relationship between the output of the dc amplifier stage and the specification part-characteristics chosen.

Prediction

Eq. (5) expressed the relationship between the output E_1 of the dc amplifier and the specification part-characteristic differences in plate currents (I_b) and differences in resistances (R). With knowledge of the distributions of these differences for the populations involved, it is possible to predict the distribution of the output by substituting into the equations

$$\mu_{E_1} = b_0 + b_1\mu_1 + b_2\mu_2$$

$$\sigma_{E_1}^2 = b_1^2\sigma_1^2 + b_2^2\sigma_2^2 + \sigma_\epsilon^2$$

Here, μ_1 and μ_2 are the means of the distributions of $(I_{b_1} - I_{b_2})$ and $(R_1 - R_2)$, respectively. Also, σ_1^2 and σ_2^2 are the variances of $(I_{b_1} - I_{b_2})$ and $(R_1 - R_2)$, and σ_ϵ^2 is the mean square of the unexplained variation, as obtained from the regression analysis.

The distribution of plate current for each tube is not directly obtainable from the applicable MIL-E-1 specification. However, it may be estimated from the specification minimum-maximum limits, employing the method suggested by Miles. Thus, the estimated distribution of I_b may be defined by

$$\mu_{I_b} = 0.7 \text{ mAdc}$$

$$\sigma_{I_b} = .067 \text{ mAdc.}$$

Similarly, from the resistor-procurement specifications requiring 3.3K ± 10 per cent resistors, the distribution of R can be estimated as

$$\mu_R = 3300 \text{ ohms}$$

$$\sigma_R = 110 \text{ ohms.}$$

From the above estimates, the distributions of the differences can be obtained.[12]

$$\mu_1 = \mu_{I_{b_1}} - \mu_{I_{b_2}} = .7 - .7 = 0 \text{ mAdc}$$

$$\sigma_1^2 = \sigma_{I_{b_1}}^2 + \sigma_{I_{b_2}}^2 = (0.67)^2 + (.067)^2 = .0089 \text{ mAdc}^2 \tag{6}$$

and

$$\mu_2 = \mu_{R_1} - \mu_{R_2} = 3300 - 3300 = 0 \text{ ohms}$$

$$\sigma_2^2 = \sigma_{R_1}^2 + \sigma_{R_2}^2 = (110)^2 + (110)^2 = 24{,}200 \text{ ohms}^2. \tag{7}$$

The distribution of the output of the dc amplifier can now be predicted by

[10] R. L. Anderson and T. A. Bancroft, "Statistical Theory in Research," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 153-190; 1952.

[11] *Ibid.*, pp. 191-206.

[12] Implicit in (6) is the assumption that I_{b_1} and I_{b_2} are distributed independently. Actually, this is not likely to be true, because the tubes V_1 and V_2 will probably be drawn from the same lot, rather than be selected at random from the entire population of type 5719 tubes. Since the covariance between I_{b_1} and I_{b_2} is positive, the effect of neglecting the term is to overestimate somewhat the size of σ_1^2. The same considerations apply to (7).

$$\mu_{E_1} = b_0 + b_1\mu_1 + b_2\mu_2$$

$$= 100.17 - 56.04(0) + .016(0)$$

$$\mu_{E_1} = 100.17 \text{ volts}$$

$$\sigma^2_{E_1} = b_1^2\sigma_1^2 + b_2\sigma_2^2 + \sigma_\epsilon^2$$

$$= (056.04)^2(.0089) + (.016)^2(24200) + 2.06$$

$$= 28.20 + 6.07 + 2.06$$

$$= 36.33$$

and

$$\sigma_{E_1} = 6.0 \text{ volts.}$$

The numerical values obtained for the regression coefficients, b_i, and for the variance, σ_ϵ^2, of the error term are estimates which differ from the true values by unknown amounts. Before any reliance can be placed on the values of μ and σ^2 calculated using these numerical values, some idea of the probable error involved must be available. A method for calculating the probable error is given in the Appendix. For the output of the dc amplifier, the confidence limits calculated for the mean and the standard deviation were ± 0.6 per cent and ± 11.0 per cent of the predicted values, respectively. If greater accuracy is desired, additional sets of measurements must be taken on the circuit.

Cathode Follower

The cathode-follower stage of the oscillator was treated in the same manner as the dc amplifier, utilizing the output, E_1, of the dc amplifier as an independent variable in the regression model

$$E_2 = b_0 + b_1E_1 + b_2I_b + b_3R_3 + \epsilon.$$

The mean and variance of E_2 were estimated by the same procedures that were used in the solution of the dc amplifier. The only difference was the requirement for knowledge of the mean and variance of E_1. However, an estimate of the distribution of E_1 was available from the solution of the dc amplifier.

The predicted distribution of the output of the dc amplifier, as obtained through the method described in this paper, is presented in Fig. 4, along with the distribution of output obtained from the experimental sample. Comparison of the two distributions serves to point out the necessity of population-distribution estimates. If the output estimates are based on a sample, they may fail to yield an accurate estimate of the population mean. Moreover, in most cases, they will be over optimistic on the σ^2 (or tightness) of the distribution. The importance of the best possible population estimation is obvious when the circuit performance characteristics of interest are critical to satisfactory circuit performance.

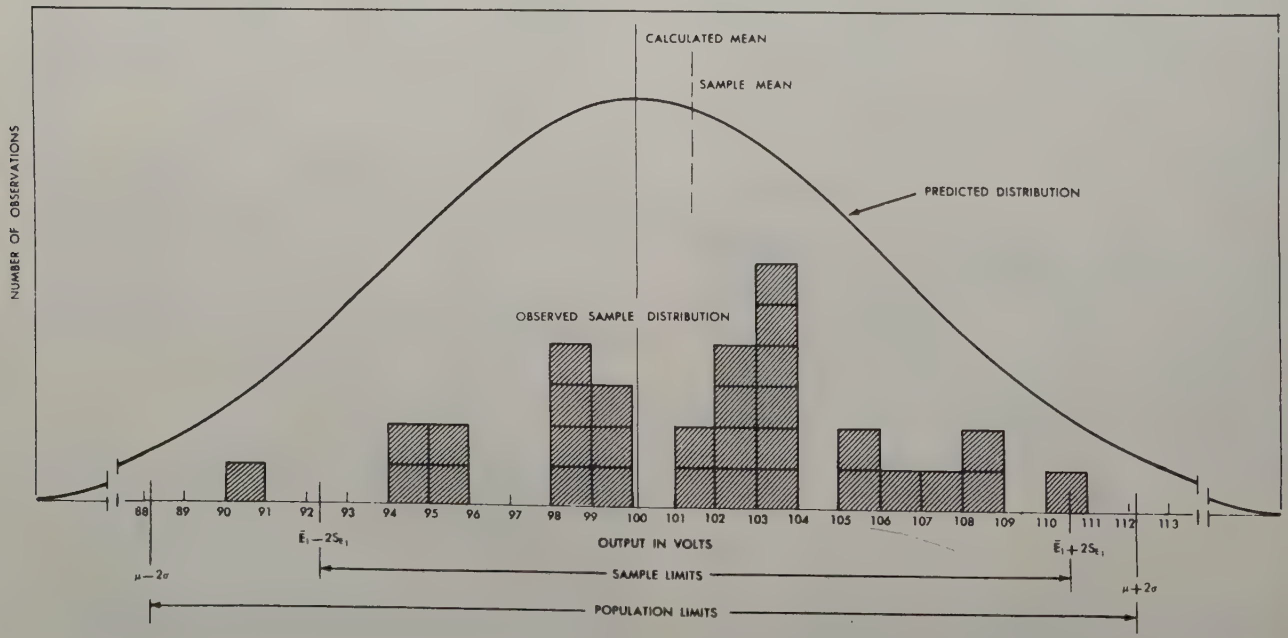


Fig. 4—Output of the dc amplifier. Histogram: measured output of experimental sample. Curve: predicted output of dc amplifier population.

Multivibrator

The model of the regression equation assumed for the multivibrator was

$$f = b_0 + b_1 E_2 + b_2 (I_{b_4} + I_{b_5}) + b_3 (I_{b_4 \text{ cut-off}} + I_{b_5 \text{ cut-off}}) + b_4 (R_5 C_1 + R_4 C_2) + \epsilon \ .$$

This form requires some explanation. The symmetry of the circuit lead to adoption of the use of the sum of part characteristics for the approximating functions. The choice of the time constant $(R_5 C_1 + R_4 C_2)$ as one of the approximating functions was dictated by the theory of operation of the circuit. As it was not known how the other terms should be selected, a linear assumption was made. Originally, the equation also contained terms involving R_6, R_7, and some additional tube characteristics. Because these terms proved not to be significant, they were dropped from the equation. Analysis of the multivibrator circuit, using the equation given above, yielded a predicted mean of 2180 cps and a predicted standard deviation of 87 cps for the output frequency.

Oscillator

The problem chosen earlier—to predict the distribution of the nominal output frequency of the oscillator—is now solved. Under the series method of solution used, the predicted output of the oscillator is the same as that of the multivibrator, given above.

The prediction obtained on the oscillator circuit provides an indication of the performance to be expected from a large number of these oscillators. This furnishes the designer a means of determining in advance whether his design is adequate. The accuracy of the prediction, of course, can be completely verified only through large-scale production of the oscillators.

IV. CONCLUSION

The described method[13] for predicting initial circuit performance consists of three distinct steps: 1) the use of regression analysis to obtain a relationship between circuit performance and part performance, 2) the use of the specifications to determine the distribution of part performance, and 3) mathematical combination of the first two steps to make a performance prediction.

Value of Regression Techniques

The application of multiple-regression techniques to circuit problems is valuable in studying the relationships between circuit performance and the various factors on which it depends. For example, if regression analysis revealed that the circuit-performance characteristic under consideration was not dependent on the specification part characteristics, additional part characteristics—not listed in the part specifications—could be included, in order to determine whether circuit performance depended upon them. Under the circumstances, inclusion of these additional characteristics in the regression equation would add nothing to the performance prediction. However, if the analysis revealed that the additional part characteristics did indeed account for the variability in circuit performance, it would be logical to conclude either that the variability not controlled by the specifications must be accepted, or that additional tests are required for parts intended for use in the circuit in question. In any event, a firm knowledge of the factors causing variability in a particular circuit would be of considerable value.

Importance of the Specifications

The method described in this paper relies on the specifications for information about the distribution of part characteristics. It is realized that present-day specifications are frequently inadequate for supplying this information. The use of the specifications, however, is nevertheless better than the alternative, which is to base the estimation on actual measurements on the parts. To base a prediction on measurements taken on a single lot of parts is never satisfactory; and, while basing the prediction on measurements taken on a large number of parts is better, it offers no real assurance for the future. If more realistic performance predictions are to be made, specifications must be changed to set forth distribution requirements on part characteristics.

[13] It should be noted that the mathematical apparatus involved in this method of prediction can, in principle, be applied to predict the distribution of any variable that is closely dependent on other variables whose distribution is known or can be predicted. For example, the output performance of an equipment or system could be predicted from a knowledge of the performance of the major subassemblies. An important consideration would be the sample size required.

Value of Performance Prediction

By means of a circuit-performance prediction, a designer can find out directly if his design will actually fall on the design-center value intended, and if the spread of the performance characteristic will be too great to tolerate. When the design is not satisfactory, the prediction not only brings the error to light, but also provides a ready means for calculating the numerical values of the changes that must be made in the nominal values and tolerances of the parts in order to obtain the required output.

APPENDIX: CONFIDENCE LIMITS

In most cases, it would be desirable to have tolerance limits on the distribution of the circuit performance characteristic, Y. If μ and σ^2 were known exactly, and if the distribution of Y were assumed to be normal, one could state that 95 per cent of the distribution of Y lies within the limits $\mu \pm 1.96\sigma$. Since μ and σ^2 can only be estimated, one would like to be able to find a number $L > 1.96$ such that the estimated limits $\hat{\mu} \pm L\hat{\sigma}$ contain 95 per cent of the distribution of Y, 95 per cent of the times that these limits are calculated. This problem has been solved in the case of samples from a normal distribution.[14] The problem here is more difficult, however, for $\hat{\mu}$ and $\hat{\sigma}^2$ are not independently distributed in general, and the distribution of $\hat{\sigma}^2$ is not a simple Chi-Square. Instead of tolerance limits on the distribution of Y, confidence limits on $\hat{\mu}$ and $\hat{\sigma}^2$ can be derived. These limits do provide some idea of the probable magnitude of the error in $\hat{\mu}$ and $\hat{\sigma}^2$ due to sampling errors in the calculated values of the $\hat{b}$'s and $\hat{\sigma}_\epsilon^2$. In addition to this sampling error, of course, the error due to estimation of the μ_i and the σ_{ij} from the specifications must be considered in any practical problem.

Confidence Limits on μ:

Confidence limits on μ are relatively easy to obtain. $\hat{\mu}$ is a linear combination of $\hat{b}_o \ldots \hat{b}_k$. Thus

$$\hat{\mu} = \hat{b}_o + \hat{b}_1\mu_1 + \ldots + \hat{b}_k\mu_k .$$

It can be shown[15] from regression theory that the variables $\hat{b}_o \ldots \hat{b}_k$ are normally distributed with means $b_o \ldots b_k$, and variances and covariances given by

$$\begin{bmatrix} \mathrm{Var}(\hat{b}_o) & \mathrm{Cov}(\hat{b}_o,\hat{b}_1) & \ldots & \mathrm{Cov}(\hat{b}_o,\hat{b}_k) \\ \mathrm{Cov}(\hat{b}_1,\hat{b}_o) & \mathrm{Var}(\hat{b}_1) & \ldots & \mathrm{Cov}(\hat{b}_1,\hat{b}_k) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(\hat{b}_k,\hat{b}_o) & \mathrm{Cov}(\hat{b}_k,\hat{b}_1) & \ldots & \mathrm{Var}(\hat{b}_k) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n} + \Sigma\Sigma\overline{X}_i\overline{X}_jC_{ij}, & -\Sigma\overline{X}_iC_{i1} & \ldots & -\Sigma\overline{X}_iC_{ik} \\ -\Sigma\overline{X}_iC_{1i} & C_{11} & \ldots & C_{1k} \\ \vdots & \vdots & & \vdots \\ -\Sigma\overline{X}_iC_{ki} & C_{k1} & \ldots & C_{kk} \end{bmatrix} \sigma_\epsilon^2 \equiv [V]\sigma_\epsilon^2$$

The matrix of the quantities C_{ij} is first found by taking the inverse of the matrix of the sums of squares and products of the deviations of the X_i about their means. The elements of the first row and column of the matrix V can then be calculated. It is also possible to calculate V directly, since V is equal to the inverse of the matrix of the sums of squares and products of the X_i, but the above form is more convenient when the analysis is performed on a desk calculator.

Since $\hat{\mu}$ is a linear combination of normally distributed variables, it follows that $\hat{\mu}$ itself is normally distributed with mean

$$\begin{aligned} E(\hat{\mu}) &= E(\hat{b}_o) + E(\hat{b}_1)\mu_1 + \ldots + E(\hat{b}_k)\mu_k \\ &= b_o + b_1\mu_1 + \ldots + b_k\mu_k \\ &= \mu , \end{aligned}$$

and variance

$$\begin{aligned} E(\hat{\mu} - \mu)^2 &= E[(\hat{b}_o - b_o) + (\hat{b}_1 - b_1)\mu_1 + \ldots \\ &\quad + (\hat{b}_k - b_k)\mu_k]^2 \\ &= \mathrm{Var}(\hat{b}_o) + \mu_1^2\mathrm{Var}(\hat{b}_1) + \ldots \\ &\quad + \mu_k^2\mathrm{Var}(\hat{b}_k) + 2\mu_1\,\mathrm{Cov}(\hat{b}_o,\hat{b}_k) + \ldots \\ &\quad + 2\mu_1\mu_2\mathrm{Cov}(\hat{b}_1,\hat{b}_2) + \ldots \\ &= [1,\mu_1 \ldots \mu_k]\begin{bmatrix} \\ V \\ \\ \end{bmatrix}\begin{bmatrix} 1 \\ \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}\sigma_\epsilon^2 \equiv Q^2\sigma_\epsilon^2 \end{aligned}$$

[14] A. Wald and J. Wolfowitz, "Tolerance limits for a normal distribution," *Ann. Math. Statistics*, vol. 17, pp. 208-215; June, 1946.

[15] Hald, *op. cit.*, pp. 638-642.

Consequently, $(\hat{\mu} - \mu)/Q\sigma_\epsilon$ is normally distributed with zero mean and unit variance.

Also, from regression theory, $S_\epsilon^2 = \Sigma(\hat{Y} - Y)^2/(n - k - 1)$, and $(n - k - 1)S_\epsilon^2/\sigma_\epsilon^2$ has a X^2 distribution with $(n - k - 1)$ degrees of freedom.

Therefore $(\hat{\mu} - \mu)/Q\sigma_\epsilon(S_\epsilon/\sigma_\epsilon)$ has a t distribution with $n - k - 1$ degrees of freedom, and confidence limits are given by

$$\hat{\mu} - tQS_\epsilon \leq \mu \leq \hat{\mu} = tQS_\epsilon. \tag{8}$$

Confidence Limits on σ^2

Exact confidence limits on σ^2 have not yet been worked out. Since the distribution of $\hat{\sigma}^2$ becomes approximately normal when the number of degrees of freedom is large and when the t values associated with $\hat{b}_i$ are large, it is possible to obtain approximate confidence limits by deriving the expression for the variance of $\hat{\sigma}^2$. The expression for this variance will contain the population parameters $b_1, \ldots b_k$, and σ_ϵ^2. An approximate value for the variance then can be calculated by substituting the estimates of these parameters into the expression, and an expression analogous to (8) will give approximate confidence limits on σ^2. The derivation is best carried out in matrix notation.

Let $B \equiv [b_1, \ldots, b_k]$ and $\hat{B} \equiv [\hat{b}_1, \ldots, \hat{b}_k]$,

$$\text{and let } S \equiv \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot\cdot & \cdot & & \cdot \\ \sigma_{k1} & \sigma_{k1} & \cdots & \sigma_{kk}^2 \end{bmatrix}$$

$$\text{and } C \equiv \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ C_{k1} & C_{k2} & \cdots & C_{kk} \end{bmatrix}$$

then,

$$\hat{\sigma}^2 = \hat{B}S\hat{B}' + S_\epsilon^2$$

where a prime is used to designate the transpose of a matrix, and S_ϵ^2 is the unbiased estimate of σ_ϵ^2.

$$\begin{aligned}\hat{\sigma}^2 &= [(\hat{B} - B) + B]S[(\hat{B} - B) + B]' + S_\epsilon^2 \\ &= BSB' + (\hat{B} - B)SB' + BS(\hat{B} - B)' \\ &\quad + (\hat{B} - B)S(\hat{B} - B)' + S_\epsilon^2 .\end{aligned}$$

If it is assumed that the t values calculated in the regression analysis are all large, then the range of $\hat{b}_i - b_i$ will be small compared to b_i. The fourth term in the equation above can then be neglected and

$$\hat{\sigma}^2 \approx BSB' + 2BS(\hat{B} - B)' + S_\epsilon^2 .$$

Taking the expected value

$$E(\hat{\sigma}^2) = BSB' + \sigma_\epsilon^2 = \sigma^2 ,$$

also,

$$E(\hat{\sigma}^2 - \sigma^2)^2 = E[2BS(\hat{B} - B)' + (S_\epsilon^2 - \sigma_\epsilon^2)]^2 ,$$

then

$$\begin{aligned}E(\hat{\sigma}^2 - \sigma^2)^2 &= E[4BS(\hat{B} - B)'(\hat{B} - B)SB' + (S_\epsilon^2 - \sigma_\epsilon^2)^2 \\ &\quad + 4(S_\epsilon^2 - \sigma_\epsilon^2)BS(B - B)'] \\ &= 4BSCSB'\sigma_\epsilon^2 + 2\sigma_\epsilon^4/f,\end{aligned}$$

since the variance[16] of S_ϵ^2 is $2\sigma_\epsilon^4/f$, and since S_ϵ^2 is distributed independently of the b_i.

Because the values of B and σ_ϵ^2 in the above expression are not known, these quantities will be replaced by their estimates. The error introduced will not be serious if the t values and degrees of freedom are large.

$$E(\hat{\sigma}^2 - \sigma^2)^2 \approx [4\hat{B}SCS\hat{B}' + 2S_\epsilon^2/f]\sigma_\epsilon^2 \equiv p^2\sigma_\epsilon^2 .$$

Approximate confidence limits on σ^2 are therefore

$$\hat{\sigma}^2 - tS_\epsilon P \leq \sigma^2 \leq \sigma^2 + tS_\epsilon P. \tag{9}$$

Application to the dc Amplifier

The estimated mean at the output of the dc amplifier is

$$\hat{\mu}_{E_1} = \hat{b}_0 + \hat{b}_1\mu_1 + \hat{b}_2\mu_2 = \hat{b}_c .$$

Q in (8) reduces to

$$Q = \sqrt{\frac{1}{n} + \overline{X}_1^2 C_{11} + 2\overline{X}_1\overline{X}_2 C_{12} + \overline{X}_2^2 C_{22}} .$$

The value of t for 28 degrees of freedom at the 0.05 level of significance is 2.05. Ninety-five per cent confidence limits on the mean are, therefore,

$$\begin{aligned}\mu_{E_1} &= \hat{\mu}_{E_1} \pm tS\epsilon Q \\ &= 100.17 \pm (2.05)(1.434)(.1982) \\ \mu_{E_1} &= 100.17 \pm .58 \text{ volts.}\end{aligned}$$

From the symmetry of the dc amplifier circuit, it is evident that the true value of μ_{E_1} is one-half of B^+, or 100 volts. The agreement is good.

The estimated variance at the output of the dc amplifier is

[16] *Ibid.*, pp. 276-278.

$$\hat{\sigma}^2_{E_1} = \hat{b}_1^2\sigma_1^2 + \hat{b}_2^2\sigma_2^2 + S_\epsilon^2 . \tag{10}$$

The expression

$$P^2 = 4\hat{B}SCS\hat{B}' + 2S_\epsilon^2/f$$

in (9) becomes

$$P^2 = 4[\hat{b}_1\hat{b}_2]\begin{bmatrix}\sigma_1^2 & O\\ O & \sigma_2^2\end{bmatrix}\begin{bmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{bmatrix}\begin{bmatrix}\sigma_1^2 & O\\ O & \sigma_2^2\end{bmatrix}\begin{bmatrix}\hat{b}_1\\ \hat{b}_2\end{bmatrix} + 2S_\epsilon^2/f$$

$$P^2 = 4(\hat{b}_1^2\sigma_1^4 C_{11} + \hat{b}_2^2\sigma_2^4 C_{22} + 2\hat{b}_1\hat{b}_2 C_{12}\sigma_1^2\sigma_2^2) + 2S_\epsilon^2/f.$$

Making use of the equations $t_{b_1} = \hat{b}_1/S_\epsilon\sqrt{C_{11}}$ and $t_{b_2} = \hat{b}_2/S_\epsilon\sqrt{C_{22}}$, P^2 becomes

$$P^2 = 4\hat{b}_1^4\sigma_1^4/S_\epsilon^2 t_{b_1}^2 + 4\hat{b}_2^4\sigma_2^4/S_\epsilon^2 t_{b_2}^2 + 8\hat{b}_1^2\hat{b}_2^2\sigma_1^2\sigma_2^2 C_{12}/t_{b_1}t_{b_2}\sqrt{C_{11}}\sqrt{C_{22}} + 2S_\epsilon^2/f.$$

Substituting for P into (9), the confidence limits on σ^2 can be written as

$$\hat{\sigma}^2 \pm [(\hat{b}_1^2\sigma_1^2 \cdot 2t/t_{b_1})^2 + (\hat{b}_2^2\sigma_2^2 \cdot 2t/t_{b_2})^2 + 2(\hat{b}_1^2\sigma_1^2 \cdot 2t/t_{b_1})(\hat{b}_2^2\sigma_2^2 \cdot 2t/t_{b_2})C_{12}/\sqrt{C_{11}}\sqrt{C_{22}} + 2S_\epsilon^4 t^2/f]^{\frac{1}{2}}.$$

In this form, it is possible to interpret the significance of the various terms in the brackets. The quantity t/t_{b_1} found in the first term is, roughly speaking, the fractional error in $\hat{b}_1$. Doubling this error gives approximately the fractional error in $\hat{b}_1$ squared. The first term, consequently represents the square of the absolute error in the first term of (10). The second term represents the square of the error in the second term of (10). If $\hat{b}_1$ and $\hat{b}_2$ were independently distributed, the errors in the first and second terms of (10) would add as the square root of the sum of the squares. The third term in the equation above, which contains the correlation coefficient $C_{12}/\sqrt{C_{11}}\sqrt{C_{22}}$ takes into account any correlation between $\hat{b}_1$ and $\hat{b}_2$, while the last term above represents the error in the estimate S_ϵ^2, which, being independently distributed, adds as the square root of the sum of the squares.

Substituting numerical values obtained in the dc amplifier problem, there results

$$\sigma^2_{E_1} = 36.3 \pm [54.5 + 13.5 - 3.8 + 1.3]^{\frac{1}{2}}$$
$$= 36.3 \pm 8.1 = 36.3 \pm 22\%.$$

Consequently,

$$\sigma_{E_1} = 6.0 \pm 11\% \text{ volts.}$$

RELIABILITY USING REDUNDANCY CONCEPTS*

L. DEPIAN† *and N. T. GRISAMORE*†

Summary—This paper introduces a new method of using redundancy to obtain reliable operation of electronic circuits. Switching circuits are used as examples to illustrate the method. A comparison is made between the majority logic method and the averaging method proposed in this paper. The comparison shows that the averaging method should provide a greater circuit reliability than the majority method if the components of each circuit have the same reliability.

INTRODUCTION

Reliability, the probability of occurrence of a desirable result, has always concerned man. As soon as he learned to control the outcome of any particular function, he started striving for higher reliabilities. In modern days, this desire is still the same; and as our world has grown more complex, higher reliabilities have become more vital.

In some cases an undesirable result or error is recognized as such; we know what the result should be and any erroneous nature is readily recognized. In other cases, however, the outcome is not known. We are seeking information and are not in a position to know the correctness of a particular outcome. Such is, for example, the nature of measurements; if the measuring instrument is defective, the result will be in error and in general the error will be unrecognized, at least in the immediate sense.

This is also the nature of the problem concerning reliable operation of systems composed of digital circuits. At present large scale digital computers use from 1,000 to 10,000 gate circuits, each circuit composed of a number of passive linear elements and one or more non-linear devices such as diodes, transistors, vacuum tubes, and/or magnetic cores. If to each of the gates we attach a number representing the probability of reliable operation of the circuit for some definite time interval, then an upper bound of the system reliability can be computed. For example, consider the case where the gate circuit is 99.999 per cent reliable, meaning the gate has a probability of 0.99,999 of operating correctly during the specified time interval. The reliability of a computer composed of 10,000 gates would then be only $(0.99,999)^{10,000}$ or about 0.905. Even if the computer used only one tenth of its gates during the specified time interval, the reliability would be only about 0.990.

To regard the problem in a different aspect, suppose the circuit in question operates at a frequency of one megacycle, which is a reasonable figure for present day computers. If the system is composed of 1000 gates, each having a reliability of 0.99,999, then the system would on the average generate errors at the rate of 10,000 per second.

Regarding the above figures, one is inclined to wonder how complex systems can be made to operate usefully and economically. In actual systems, however, only a few component parts operate continuously, thus reducing the probability of failure of the whole system. Secondly, in some cases, an error produced by one of the gates will not be propagated throughout the whole system. Regardless of these factors, however, the reliable operation of large scale systems is still a problem, as evidenced by the amount of effort expended in increasing component reliability, devising error-checking schemes, marginal checking, preventive maintenance, etc.

NATURE OF THE PROBLEM

In general, computer and logical circuits, consisting of gates, are designed to represent the function

$$A = F(a,b,c,------), \tag{1}$$

where A is the output, $a,b,c,------$ are the inputs, and $F(a,b,c,---)$ is some particular Boolean function of the inputs. Since this is a Boolean equation, the variables can assume only two possible values corresponding to two possible states of the circuit which we will call "on" and "off".

The problem of reliable operation is one where we wish the circuit shown in Fig. 1 to represent function (1) with a given probability. In general, by the simple process of repetition (redundancy)

*This work was supported by the Office of Naval Research, Washington, D. C., under Contract No. N7on41906.

†Elec. Engrg. Dept., The George Washington University, Washington, D. C.

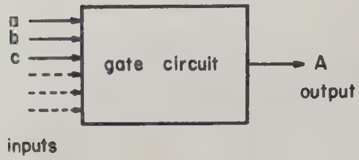


Fig. 1—Generalized gate circuit represented by (1).

one can increase the reliability of a system. For instance, if there is a number of circuits, all the same, in parallel, a failure of one will not be catastrophic since the others will still carry on their function. The difficulty here, however, lies in the fact that outputs and inputs may not always be in a state labeled "on" or "off". As an example, suppose an output of 2 ma is labeled as "off" and one of 10 ma as "on". Because of malfunctions, however, (defective components, spread and/or drift in characteristics, aging, etc.) the output may be 5 ma, neither "on" nor "off". A dividing line may be thought of, for example, at 5.5 ma and any output smaller than this value labeled as "off" and any larger output as "on". However, even this scheme is not satisfactory since there may still be "off" outputs larger than 5.5 ma giving rise to an error, since they will be considered as "on" by any device sensing the output. In general, a probability density will exist for the "off" and "on" outputs, as shown in Fig. 2. It is assumed that this distribution takes into account any effect caused by a distribution of amplitudes of the input signals. An output to the right of the dividing line will be considered as "on," and to the left as "off." Clearly, the area A will give the probability of error in the "off" state and area B the probability in the "on" state. The problem is to decrease these error probabilities.

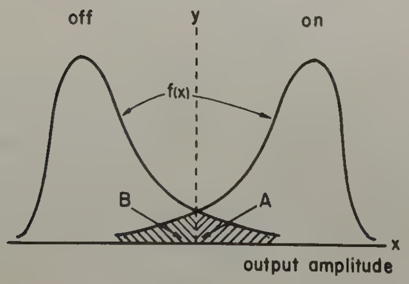


Fig. 2—Probability density, $f(x)$, for outputs of amplitude x from gate circuits of the type shown in Fig. 1.

THE MAJORITY METHOD

Recent efforts have been made to apply the principle of redundancy to large-scale systems, so as to increase the reliability of these systems. These efforts have been based on ideas proposed by von Neumann.[1] The scheme is to use identical redundant circuits, each having the same input signals, and one majority device which senses the outputs of the redundant circuits. The majority device then produces an output which is in agreement with the majority of the outputs. For example, consider three redundant circuits. If all three outputs fall on one side of the dividing line in Fig. 2, let us say to the left, or two on the left and one on the right, the output will be considered by the majority device as "off." Let us see what improvement has been obtained in this fashion. Consider the "off" probability density $f_1(x)$ of the output x, and let y be the value of x at the dividing line. The probability of correct interpretation using one circuit will depend on the position of y and will be R_1 (the probability of x less than y), shown in Fig. 3.

$$R_1 = \int_{-\infty}^{y} f_1(x)dx. \tag{2}$$

If three circuits are used in the majority scheme, a correct interpretation will be made 1) if all three circuit outputs are less than y giving: probability = R_1^3; or if 2) two of the outputs are less than y and one is greater than y giving: probability = $3R_1^2(1 - R_1)$.

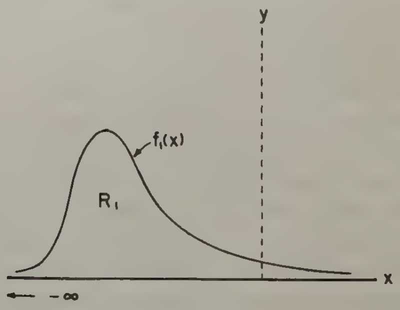


Fig. 3—Reliability, R_1, as a function of the distribution of the output amplitude, x. R_1, represents the area under curve between the limits -oo and y.

[1] J. von Neumann, "Probabilistic Logics and Synthesis of Reliable Organisms from Unreliable Components," Automata Studies, Princeton University Press, Princeton, N. J.; 1956.

The new probability of correct interpretation will then be

$$R_{3,m} = R_1^2(3 - 2R_1) \quad (3)$$

which is an improvement over R_1 if R_1 is greater than 0.5.

If five circuits are used in a majority circuit, the new probability is found to be

$$R_{5,m} = R_1^3(10 - 15R_1 + 6R_1^2) \quad (4)$$

and for a majority of seven circuits

$$R_{7,m} = R_1^4(35 - 84R_1 + 70R_1^2 - 20R_1^3) . \quad (5)$$

These functions are shown in Fig. 4.

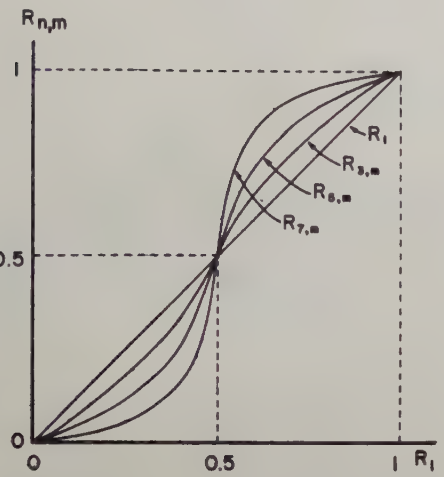


Fig. 4—Reliability, $R_{n,m}$, of majority method compared with reliability of single circuit, R_1.

It is seen that, if the original probability R_1 of obtaining a correct interpretation of the output for a single circuit is larger than 0.5, the majority gives a definite improvement in reliability, assuming, of course, a perfect majority device. Furthermore, this improvement is increased as the number of redundant circuits is increased.

The majority method, however, makes no use of the probability density curve of the circuit output. R_1 can have the same value for a different $f_1(x)$ and a different y. $R_{n,m}$ is derived directly from R_1 and in that sense is independent of the form of $f_1(x)$. In other words, the majority method does not make full use of the information associated with $f_1(x)$. It could be asked at this point: is the majority method the best reliability improvement? Could it not be that a different method, possibly making use of the nature of $f_1(x)$, would give better reliabilities?

THE AVERAGING METHOD

The reliability associated with the "off" state could be improved by increasing y (see Fig. 2), but this would be done at the expense of decreasing the reliability of the "on" state. The value of y is to some extent fixed and cannot be used directly to improve the over-all reliability.

Returning to Fig. 3, it may be seen that the reliability would increase if the probability density $f_1(x)$ could be adjusted so that less of the area under the curve lies to the right of y. Suppose that n redundant circuits are to be used to increase the reliability, and that their outputs are $x_1, x_2, x_3, \ldots, x_n$, each following the same probability density $f_1(x)$. Let

$$s = g(x_1, x_2, x_3, \ldots, x_n) \quad (6)$$

be a function of these outputs. s will follow a probability density $f_n(s)$ which will be different than $f_1(x)$ (see Fig. 5). The reliability (probability that s is less than y) will be

$$R_{n,g} = \int_{-\infty}^{y} f_n(s)ds. \quad (7)$$

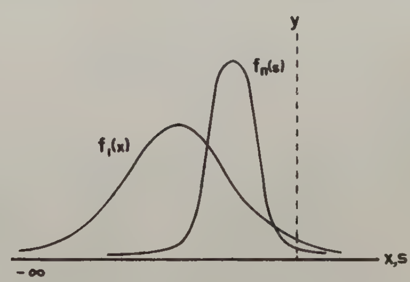


Fig. 5—Desired distribution function of s compared with x for the "off" state.

It is conceivable that one might now have

$$R_{n,g} > R_{n,m} \quad (8)$$

where $R_{n,m}$ is the reliability obtained with n redundant circuits, by the majority method.

The problem is centered in (6): can a function g be found which satisfies inequality (8)? A generalization of this would be: can one find a function g such that inequality (8) is maximized? An

attempt will be made to answer the first question by using a particular function g. The answer to the second question is discussed at the conclusion of this paper.

Consider the function

$$s = (x_1 + x_2 + x_3 + \ldots + x_n)/n. \tag{9}$$

The x's, being the outputs of the redundant circuits, will in general be independent and will each obey the same probability density curve, $f_1(x)$.

A general method of finding the probability density, $f_n(s)$, of s is by use of the Fourier transform. The Fourier transform of $f_1(x)$ is

$$\phi_1(v) = \int_{-\infty}^{+\infty} f_1(x)e^{jvx}dx. \tag{10}$$

It can be shown that the Fourier transform of the average s is related to $\phi_1(v)$ by[2]

$$\phi_n(v) = [\phi_1(v/n)]^n \tag{11}$$

from which, by using the inverse Fourier transform, the probability density $f_n(s)$ of s may be found.

$$f_n(s) = (1/2\pi) \int_{-\infty}^{+\infty} \phi_n(v)e^{-jsv}dv. \tag{12}$$

A few examples will be examined to establish the merit of this averaging method.

Normal Distribution Let x follow a normal distribution density with a mean at $x = \mu_1$ and a standard deviation σ_1, giving

$$f_1(x) = (1/\sigma_1\sqrt{2\pi})\, e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \tag{13}$$

Eq. (10) gives

$$\phi_1(v) = (1/\sigma_1\sqrt{2\pi}) \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} e^{jux}\, dx$$

or

$$\phi_1(v) = e^{-(\sigma_1 v)^2/2} e^{jv\mu_1} \tag{14}$$

from which one gets

$$\phi_n(v) = e^{-(\sigma_1 v)^2/2n} e^{jv\mu_1} \tag{15}$$

Applying the inverse Fourier transform gives

$$f_n(s) = (1/2\pi) \int_{-\infty}^{+\infty} e^{-(\sigma_1 v)^2/2n} e^{jv\mu_1} e^{-jsv}\, dv$$

or

$$f_n(s) = (1/\sigma_n\sqrt{2\pi})\, e^{-\frac{(s-\mu_1)^2}{2\sigma_n^2}} \tag{16}$$

where

$$\sigma_n = \sigma_1/\sqrt{n}\ . \tag{17}$$

The average s will also obey a normal distribution curve with the same mean but with a smaller standard deviation (see Fig. 6). The increase in reliability is evident. Let us now

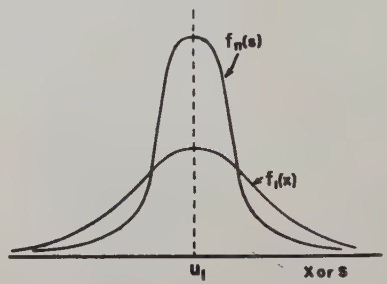


Fig. 6—Comparison of distributions of x and s, where s is the average value of x.

compare this with the majority method for $n = 3$. Let $R_1 = 0.90$. A majority system will give $R_{3,m} = 0.972$. The average will give $R_{3,a} = 0.987$. Fig. 7 gives a direct comparison between the majority method $(R_{n,m})$ and the averaging method $(R_{n,a})$ for $n = 3$ and $n = 5$ respectively. It can be seen that for R_1 (reliability of a single circuit) larger than 0.5, the averaging method will always give better results than the majority method. For R_1 less than 0.5, both methods will give reliabilities smaller than R_1.

This example dealt with an $f_1(x)$ following a normal distribution. This, of course, is not a very realistic distribution, since the output x is limited, by the physical nature of the circuit, to an upper and lower bound. In the following example we consider the more realistic case.

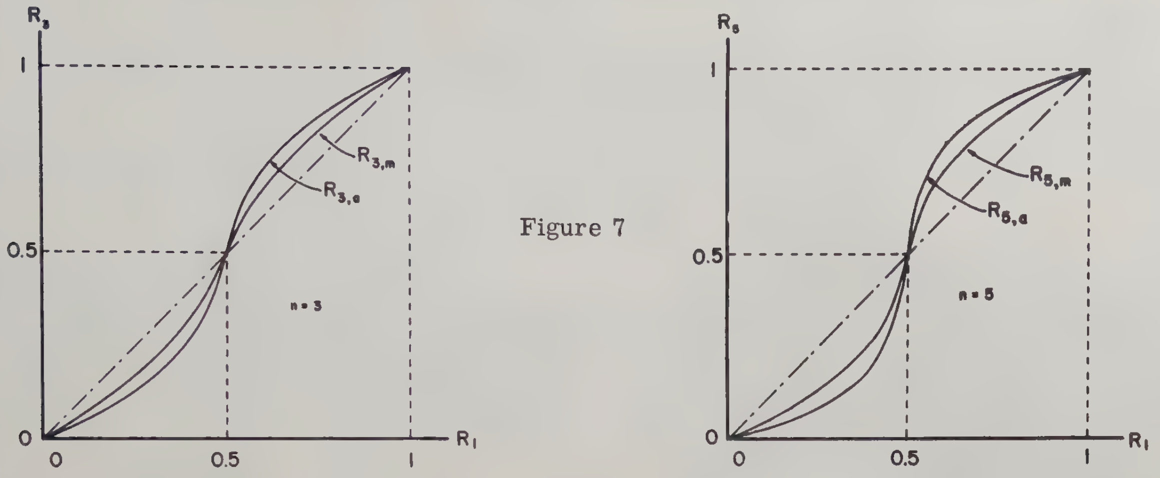
Pearson Distribution Let $f_1(x)$ be represented by a Pearson's curve of type I, where α_1 and α_2 are the lower and upper limits respectively.

$$f_1(x) = (\alpha_2 - \alpha_1)^{-(p_1+q_1)}[(p_1 + q_1 - 1)!/(p_1 - 1)!\,(q_1 - 1)!](x - \alpha_1)^{p_1 - 1}(\alpha_2 - x)^{q_1 - 1}. \tag{18}$$

The general shape of (18) is shown in Fig. 8.

Such a curve is more realistic than the normal

[2] J. V. Uspensky, "Introduction to Mathematical Probability," McGraw-Hill Co., Inc., New York, N. Y.; 1937.



(a) (b)

Fig. 7—Comparison of reliabilities obtained by majority method and average method for two different values of redundancy.

distribution and was found to give a good fit for transistor gate circuits. The sharpness of the curve depends on the form factors, p_1 and q_1. The coefficient

$$(\alpha_2 - \alpha_1)^{-(p_1 + q_1 - 1)}[(p_1 + q_1 - 1)!/(p_1 - 1)!(q_1 - 1)!]$$

normalizes the curve so that

$$\int_{\alpha_1}^{\alpha_2} f_1(x)\, dx = 1. \tag{19}$$

This imposes a lower bound on p_1 and q_1: p_1 and q_1 must be equal to or greater than 0. The mean occurs at $x = \mu_1$, where

$$\mu_1 = \int_{\alpha_1}^{\alpha_2} x f_1(x)\, dx = (q_1\alpha_1 + p_1\alpha_2)/(p_1 + q_1). \tag{20}$$

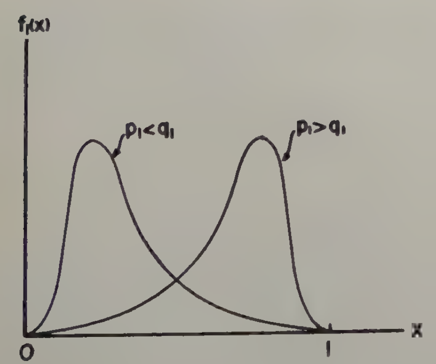


Fig. 8—Two examples of Pearson's Type I distribution curve with limits normalized to 0 and 1.

The standard deviation σ_1 is

$$\sigma_1 = \left[\int_{\alpha_1}^{\alpha_2} (x - \mu_1)^2 f_1(x) dx\right]^{\frac{1}{2}} = [(p_1 q_1)/(p_1 + q_1 - 1)]^{\frac{1}{2}}$$
$$(\alpha_2 - \alpha_1)/(p_1 + q_1) \,. \tag{21}$$

If $f_1(x)$ is normalized and the x-coordinate is adjusted so that the limits α_1 and α_2 become 0 and 1 respectively, (18) becomes

$$f_1(x) = [(p_1 + q_1 - 1)!/(p_1 - 1)!(q_1 - 1)!]\, x^{p_1 - 1}$$
$$(1 - x)^{q_1 - 1}. \tag{18a}$$

This normalization does not effect the values of p_1 and q_1, but the mean and standard deviation become,

$$\mu_1 = p_1/(p_1 + q_1) \tag{20a}$$
$$\sigma_1 = [(p_1 q_1)/(p_1 + q_1 - 1)]^{\frac{1}{2}}[1/(p_1 + q_1)]. \tag{21a}$$

By following the process outlined by (10) through (12), it is found that the distribution density of s (the average of the outputs of n redundant circuits) is again a Pearson's type I curve.

$$f_n(s) = [(p_n + q_n - 1)!/(p_n - 1)!(q_n - 1)!]\, s^{p_n - 1}$$
$$(1 - s)^{q_n - 1} \tag{22}$$

where

$$p_n = p_1[n + (n - 1)/(p_1 + q_1)] \tag{23}$$

and

$$q_n = q_1[n + (n - 1)/(p_1 + q_1)] \,. \tag{24}$$

The new mean μ_n is the same as μ_1,

$$\mu_n = p_n/(p_n + q_n) = p_1/(p_1 + q_1) = \mu_1 \quad (25)$$

while the standard deviation is reduced by a factor of $1/\sqrt{n}$:

$$\sigma_n = [(p_n q_n)/(p_n + q_n + 1)]^{\frac{1}{2}}[1/(p_n + q_n)] = \sigma_1/\sqrt{n}. \quad (26)$$

Moreover, the lower and upper limits 0 and 1 (α_1 and α_2 in the non-normalized coordinate) have remained the same. This result, conserving the mean while reducing the standard deviation, was also found for the case of the normal distribution.[2] The decrease of the standard deviation points to a sharper distribution density and in general higher reliabilities are to be expected.

A comparison with the majority scheme will now be made. Let us first examine the special case where $f_1(x)$ is a constant for one of the distributions. This is not very realistic, but it offers an indication of the advantages of the averaging over the majority method. For this distribution, we must have

$$f_1(x) = 1, \quad (27)$$

$$p_1 = q_1 = 1. \quad (27a)$$

The probability (reliability) of x less than y is

$$R'_1 = \int_0^y f_1(x)\,dx = y. \quad (28)$$

If three ($n = 3$) redundant circuits are used, the distribution of the average will be

$$f_3(s) = [(7!)/(3!3!)]\, s^3(1 - s)^3$$

which results from the fact that

$$p_3 = q_3 = 4.$$

This gives

$$f_3(s) = 140\, s^3(1 - s)^3 \quad (29)$$

and the new probability (reliability) of s less than y will be

$$R'_{3,a} = \int_0^y f_3(s)\,ds = R'^4_1(35 - 84R_1 + 70R_1^2 - 20R_1^3). \quad (30)$$

Comparing this with (5) we see that, for this case,

$$R'_{3,a} = R_{7,m}. \quad (31)$$

In other words, for the case of p_1 and q_1 equal to 1, the averaging method with three redundant circuits gives the same result as the majority method with seven redundant circuits. It is easy to show that, in general, for $p_1 = q_1 = 1$,

$$R'_{n,a} = R_{(3n-2),m}, \quad (32)$$

and since for n greater than 1, $3n - 2$ is greater than n, thus

$$R'_{n,a} > R_{n,m}. \quad (33)$$

Eq. (32) may also be written as

$$R_{n,m} = R'_{(n+2)/3,a} \quad (32a)$$

which states that the majority method with n redundant circuits will give the same reliability as the averaging method with $(n + 2)/3$ redundant circuits and p_1 and q_1 equal to 1. In other words,

$$R_{n,m} = \frac{n!}{\left(\frac{n-1}{2}!\right)^2} \int_0^{R_1} [x(1-x)]^{\frac{n-1}{2}}\,dx. \quad (34)$$

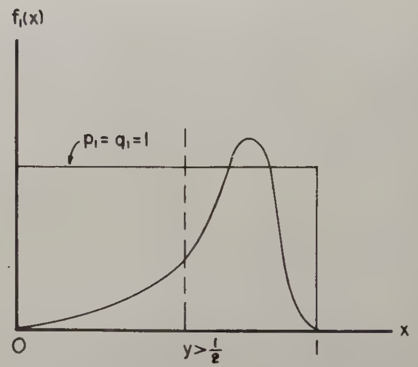


Fig. 9—Special case of distribution curve, used for comparison between averaging and majority methods. Curve on the right shows possible distribution of "on" signals that might occur.

If p_1 and q_1 are larger than unity, the averaging method will give still higher reliabilities

$$R_{n,a} > R'_{n,a}. \quad (35)$$

We conclude that

$$R_{n,a} > R_{n,m}. \quad (36)$$

Let us consider the case where $p_1 \neq q_1$. Since the choice of the labeling of "off" and "on" is arbitrary, we may always state that one would correspond to $p_1 > q_1$ and the other to $p_1 < q_1$ (see Fig. 10). Furthermore, if the state with $p_1 > q_1$ is normalized to the same coordinate as the one with $p_1 < q_1$ (shifting of upper and lower

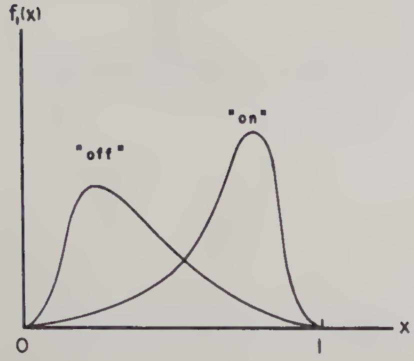


Fig. 10—Representative normalized distribution.

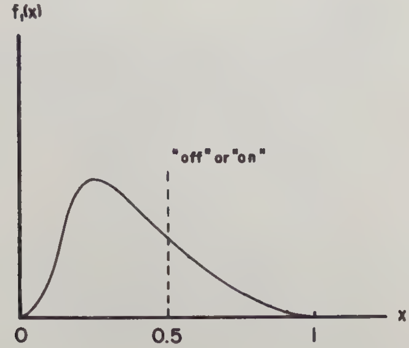


Fig. 12—Identical "on" and "off" distributions used for example in text.

limits), we may consider both "off" and "on" states with $p_1 < q_1$ (see Fig. 11). This has actually been found to be the case for transistor gates.

For simplicity, let us assume that, when normalized to the same coordinate, both "off" and "on" states present the same p_1 and q_1. This suggests that the dividing line $x = y$ should be placed at $y = 0.5$ as in Fig. 12. The single circuit reliability is

$$R_1 = \frac{(p_1 + q_1 - 1)!}{(p_1 - 1)!(q_1 - 1)!} \int_0^{0.5} x^{p_1 - 1}(1 - x)^{q_1 - 1}\, dx. \quad (37)$$

The averaging method with n redundant circuits will yield a reliability

$$R_{n,a} = \frac{(p_n + q_n - 1)!}{(p_n - 1)!(q_n - 1)!} \int_0^{0.5} s^{p_n - 1}(1 - s)^{q_n - 1}\, ds \quad (38)$$

where p_n and q_n are given by (23) and (24). The majority method with n redundant circuits will yield a reliability $R_{n,m}$, given by (34).

Figs. 13 and 14 demonstrate the advantage of the averaging over the majority method for $p_1 < q_1$ and three $(n = 3)$ redundant circuits. Fig. 14 compares the number of errors per 1,000 operations of a single circuit with those to be expected from the majority and the averaging methods. For example, for $p_1 = 3$ and $q_1 = 8.9$, a single circuit will give an erroneous output about 30 times in 1,000 operations. The majority method would give about four erroneous outputs and the averaging about one. Fig. 15 gives the ratio of the number of erroneous outputs by majority to the number by averaging. For example, for $p_1 = 2.8$ and $q_1 = 9$, the majority method

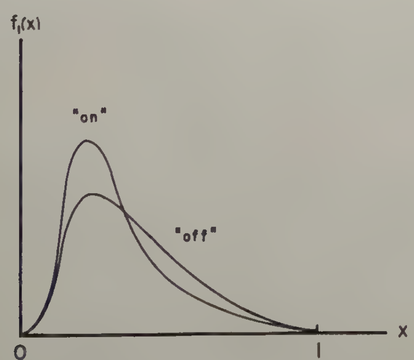


Fig. 11—"On" curve normalized to same limits as "off" curve.

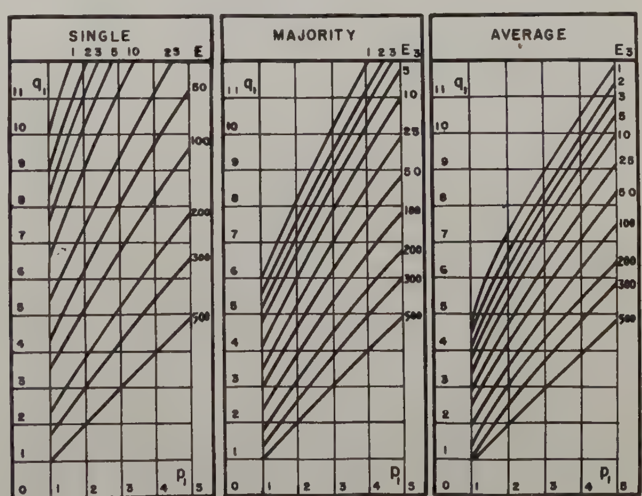


Fig. 13—Expected errors per thousand operations: a) single circuit; b) majority method with n of 3; c) averaging method with n of 3.

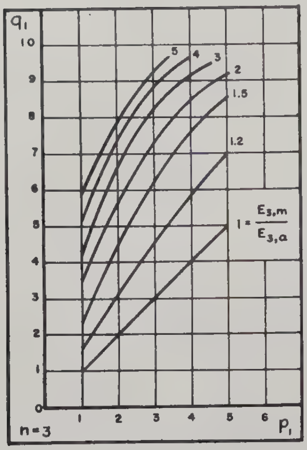


Fig. 14—Ratio of errors expected per thousand operations using the majority method, to those expected for the averaging method.

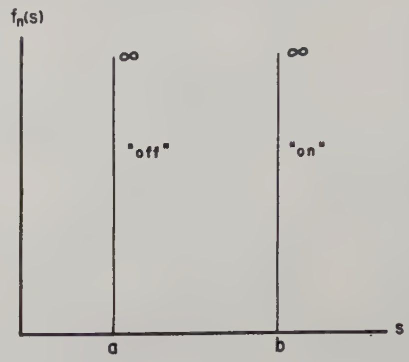


Fig. 15—Limits of distribution function, $f_n(s)$, as the redundancy is increased.

will produce five times as many errors as the averaging method.

CONCLUSIONS

It has been demonstrated that the reliability of a system may be increased by the use of redundant elements. This makes possible the use of inexpensive components, known to be less reliable, in place of more reliable and more expensive ones now used. Furthermore, it has been shown that redundancy by the averaging method will generally give better reliability improvement than the majority method. In addition, it is usually simpler to construct an averaging circuit (*e.g.*, output currents through a common load) than a majority circuit.

Although the averaging method will yield good results, it is still probably not the best. The goal would be to find functions s, (6), such that the distribution densities of the "off" and "on" states approach (as n is increased) the two Dirac functions $\delta(s - a)$ and $\delta(s - b)$. (See Fig. 15.)

The particular function s to be chosen will, in general, depend on the distribution density of the particular single circuit to be used. For example, in some cases a weighted average might prove advantageous, the weighting factor depending on the nature of $f_1(x)$.

ACKNOWLEDGMENT

The authors would like to acknowledge the aid of Mrs. Louis Depian and of G. Uyehara in the preparation of this manuscript, and of L. Rubin in calculations required to construct the graphs shown in Figs. 14 and 15.

HOW CAN WE ATTAIN HIGH RELIABILITY OF COMPLEX MILITARY ELECTRONIC EQUIPMENT?*

MORRIS HALIO†

Summary—Each piece of military electronic equipment passes through various phases in its normal life cycle. These are planning, design and development, pilot production, manufacture, transportation, storage, operation and maintenance. Each of these stages is replete with opportunities for the introduction of unreliabilities. This paper points out the pitfalls which may be encountered and makes specific recommendations to avoid these so that the full amount of potential reliability may be realized in the final equipment.

By this time, most of us have been made aware of the growing complexity of weapon systems utilizing countless electronic circuits with myriads of parts and the terrifying reliability problems that arise as a consequence thereof. Therefore, let us proceed immediately to the crux of the matter—namely, what can we do to remedy the situation? The reliability problem with its many facets is reminiscent of the many-headed hydra. Each of these heads must be removed to conquer the beast. If we were to trace a piece of equipment through its life cycle, we might arrive at a flow chart such as that shown in Fig. 1, including

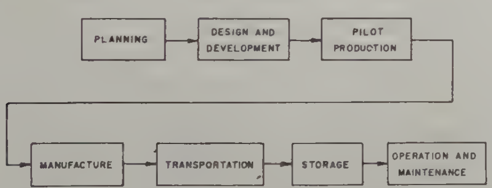


Fig. 1—Flow chart in life cycle of equipment.

such factors as: planning, design and development, pilot production, manufacturing, transportation, storage, operation and maintenance. Each of these stages presents an opportunity for additional unreliabilities to be introduced. It is obvious that if the design is such as to limit the maximum potential reliability to a certain value, then poor manufacturing processes or the deleterious effects of storage, for example, can only serve to reduce the ultimate reliability of the equipment. It is therefore imperative to minimize the unreliabilities introduced by each step in the process.

At this point, some of the terms used in this paper can be defined. The first one, of course, is "reliability." Definitions of this term vary from some long and complicated ones to the simple one, "When you press the button, it goes." This author prefers the definition employed by one of the task groups of AGREE (Advisory Group on Reliability of Electronic Equipment of the Office of the Assistant Secretary of Defense). This is: "Reliability of an item is the probability that it will perform without failure a specified function under specified test conditions for a required period of time." Incidentally, the various task groups of AGREE did not all agree on a definition for this term.

Mathematically, $R(t) = e^{-\frac{t}{m}}$, where $R(t)$ is the reliability, t is the variable time and m is referred to as the reliability index. The latter is defined as the average measure of the equipment failure rate expressed in mean-time-between-failures. The reciprocal of this quantity is known as the failure rate and is most conveniently expressed as number of failures per thousand hours.

Fig. 2 depicts a typical statistical curve of the variation of failure rate during the life of an equipment. The high rate of early failures is attributable to poor parts control, manufacturing techniques, inspection and quality control. At time A, the defective parts have been eliminated and the failure rate is essentially constant until time B, when the failure rate begins to increase, signifying the end of useful life of the equipment.

The terms employed for the various subdivisions of an equipment are still not fully standardized; therefore, this author would like to recommend the following definitions which are modifications of those listed in DOD Directive 3232.2.

*This paper was presented at the Fifth Joint Military-Industry Symp. on Guided Missile Reliability, Chicago, Ill., December 8-10, 1958.

†Headquarters Air Defense Command, USAF, Colorado Springs, Colo.; formerly at Ballistic Res. Labs., Aberdeen Proving Ground, Md.

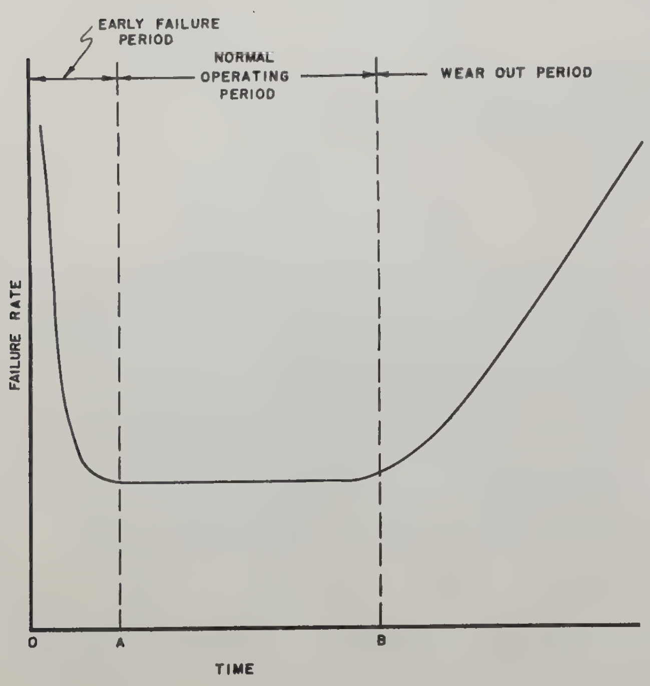


Fig. 2—Failure rate of equipment *vs* time.

Part: An item which cannot be disassembled without destroying its identity; *e.g.*, resistor, capacitor, switch, relay, socket, bearing, bolt.

Subassembly: An aggregation of parts mounted together for convenience and incapable of performing any function prior to being incorporated into an assembly; *e.g.*, a terminal board with parts mounted on it, an IF transformer with tuning slug and mechanism.

Assembly: A combination of parts or subassemblies or both capable of performing a function; *e.g.*, amplifier, oscillator, modulator, filter, power supply, junction box.

Component: An aggregation of assemblies, constituting an element of an equipment and performing a function necessary to the operation of that equipment; *e.g.*, transmitter, receiver, rotating antenna, frequency standard.

Equipment: A group of components capable of performing a specified function; *e.g.*, a radar set, a gun director.

You will notice that the subdivision formerly known as "component" is now referred to as "part," while the term "component" is reserved for designating a group of assemblies.

PLANNING

The first phase of the reliability program is the planning stage. It is necessary that quantitative specifications for equipment reliability be incorporated into the development contract. The present low level of reliability may be partly ascribed to failure to do so. In the past, a manufacturer who designed a new system has had to meet certain performance specifications. However, he has been under no legal obligation to include reliability among these. As a result, reliability has been treated as an afterthought. Long experience has shown that this is too late to improve reliability. Once the design has been frozen, the failure rate of electronic equipment cannot be appreciably decreased by debugging. High reliability cannot be achieved unless this factor is taken into account during the preceding stages.

Reliability requirements should originate with the groups responsible for the operational requirements and military characteristics of the various services, since it is through these groups that the services must determine how they intend to accomplish their mission. These figures must then be incorporated into the development contracts for new equipment. Proper planning is the foundation on which the reliability structure is based.

DESIGN AND DEVELOPMENT

Design and development follow the planning stage. The reliability of the completed equipment will depend on that of the parts employed as well as the circuitry in which they are utilized. It is well known that the over-all reliability of an equipment where the parts are placed in series[1] can be expressed by

$$R \text{ over-all} = R_1 \times R_2 \times R_3 \times \ldots \times R_n$$

i.e., the over-all reliability is the product of the individual reliabilities. The simplifying assumption has been made that there are no reliability interactions among the various parts. Evidently, for very complex equipments, the reliabilities of individual parts must be extremely high if the over-all reliability is to be tolerable. Fig. 3, which has been adopted from Lusser [3], shows the relation of over-all reliability to individual reliabilities for various degrees of equipment complexity. For simplicity, the individual reliabilities have been made equal. Notice, for example, that an equipment of 400 parts, each 99 per cent reliable, only has a 2 per cent over-all reliability. This emphasizes what is probably the most im-

[1] A part is a series part if its failure would cause the entire equipment to fail. It is a parallel part if its failure would not necessarily lead to failure of the equipment, since it is shunted by another part.

portant concept in the study of reliability, namely, that individual parts of a complex equipment must be of the very highest reliability. This means that the margins between the strengths and stresses must be sufficiently large. By stresses, we are referring not only to mechanical forces, but to other parameters such as voltage, current, frequency, temperature, humidity, acceleration, vibration, etc., to which a part is subjected in use. The strengths are the values of these parameters at which failure will occur under the given conditions.

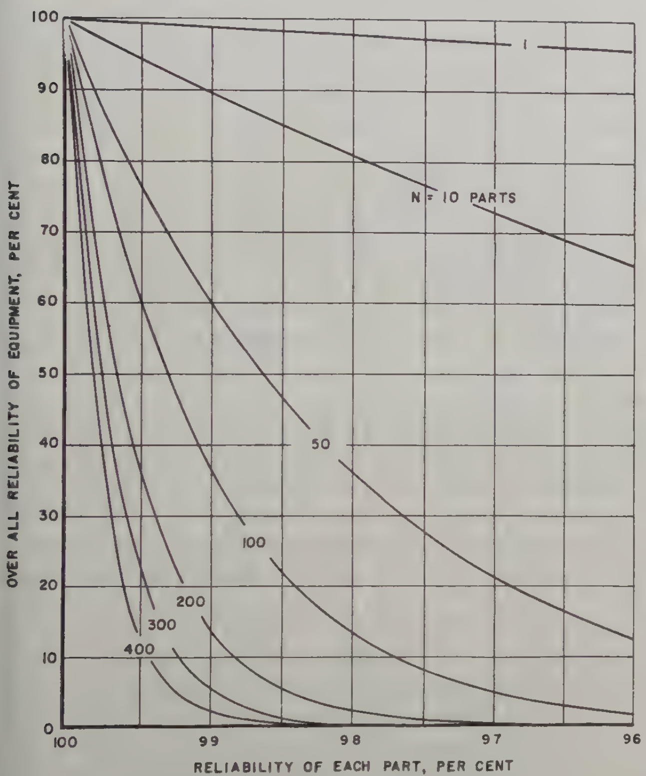


Fig. 3—Over-all reliability as a function of complexity and reliability of parts.

To determine whether a part to be used in an equipment is of acceptable reliability, a stress-strength analysis is recommended. The following procedure is employed. A stress scatter diagram is constructed as in Fig. 4, depicting the stresses to which the part will be subjected in the intended application. These data will have been obtained from field measurements. A frequency distribution curve is drawn and the mean and standard deviation calculated. Tests-to-failure are then conducted on a representative sample of the part whose use is contemplated and the strengths plotted. The frequency distribution curve, the mean and the standard deviation for the strengths are obtained. To determine the allowable margin between the mean stress and mean strength, the standard deviations of stress and strength are multiplied by suitable factors depending on the required part reliability and the products are added.

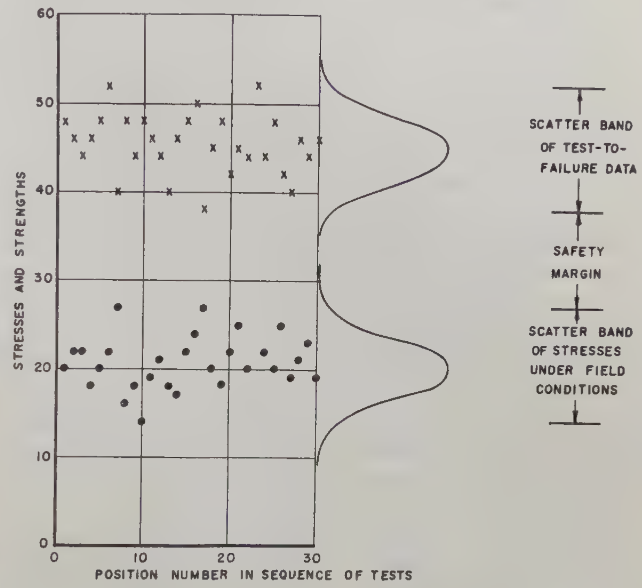


Fig. 4—Distribution curves of strengths and stresses.

Thus, the total permissible margin between the strength and stress means can be expressed as:

$$M = K_1 S_1 + K_2 S_2$$

where M is the margin, K_1 and K_2 are the strength and stress factors; S_1 and S_2 being the corresponding standard deviations. If the actual margin is less than the permissible margin, it means that the part will have to be redesigned or replaced with a more reliable part. Stress-strength analysis of this type is of extreme importance in the effort to attain high reliability.

It would be extremely desirable to standardize parts of high strength and to have this information on these parts assembled in handbooks available to designers. A start has been made in this direction, but the trend will have to be greatly accelerated to meet the needs of the military. Vitro Corporation, RCA and Inland Testing Laboratories are among those who are doing pioneering work in this field.

Sub-assemblies, assemblies and components should also be subjected to tests-to-failure. However, the purpose of testing these is to discover failures caused by specific assembly effects,

such as local resonances and ambient temperatures. Therefore testing of these items is recommended only after it has been determined that parts of extremely high reliability have been employed under conditions of adequate margin of safety. Otherwise, this type of testing becomes very cumbersome and failures attributable to part unreliability mask those caused by assembly effects.

Stress-strength analysis depends upon testing of parts to failure as contrasted to testing of complete equipments under operating conditions. Unfortunately there has been too much reliance on the latter procedure as a means of seeking the achievement of reliability. This is to be deplored, since testing-to-failure furnishes a much better means of attaining this goal. For one thing, it makes it possible to determine very quickly the modes of failure and permit redesign so that reliable parts can be used in the equipment. The old bug-hunting methods depending on failure reporting of equipments tested under normal operating conditions would take forever and a day to accomplish the desired result. In addition, testing-to-failure is far cheaper, since this method requires substantially fewer tests. Extensive flight testing of missiles, for example, can get to be rather expensive. Even then, the ultimate cause of failure is often not discovered. Or to put it another way, testing-to-failure means that we can buy much more reliability for a fixed amount of money.

One of the ways in which part reliability may be improved is for the parts designers to refrain from designing universal parts. Design of a single part for several applications with widely differing specifications tends to make the reliability for each application lower than if a different type were built for each of these. A part is generally designed for universality of application for two reasons. One is lower cost because of high quantity of production. The other is that the control processes that accompany mass production tend to improve the quality and consequently the reliability of the product. However, there is a certain level of production beyond which the quality remains essentially constant. Once this is reached, the faults of the multiplicity of functions of a universal part become evident; *e.g.*, in the case of electron tubes, a tube may be used in a dc amplifier or in a blocking oscillator. Clearly, the specifications are different for these applications. By designing a tube which is applicable to both of these uses, the reliability for each suffers. There is certainly sufficient demand for each type so that a different tube can be built for each application. It is therefore recommended that parts designers originate different types for widely varying uses.

Another step the circuit designer can take to maximize reliability is to select part types which have higher inherent reliability. Semiconductors can be used in place of electron tubes, vacuum relays instead of other types, etc.

The total effect of parts tolerances plus drift due to aging may cause failure of a circuit in operation; e.g., an oscillator may shift its frequency out of tolerance or may stop oscillating entirely; a flip-flop may reach such a condition that a prescribed pulse may fail to trigger it. To prevent such an occurrence, a design method known as marginal checking (developed by Lincoln Laboratory) is recommended. In this, the allowable variation of a part is determined as a function of a selected circuit parameter, usually a supply voltage. In practice, the tolerance of one of the parts in the circuit is plotted against the variation in this marginal-checking parameter, as shown in Fig. 5. For various values of part deviation, the supply voltage is varied until the circuit fails to perform according to specifications. The locus of failure points separates the failure region from that of normal operation. In this manner, not only can the proper design center value be determined but the allowable tolerances as well. Universal employment of marginal checking by equipment designers is decidedly recommended.

The foregoing discussion has been concerned with the reliability of parts; some of the principles

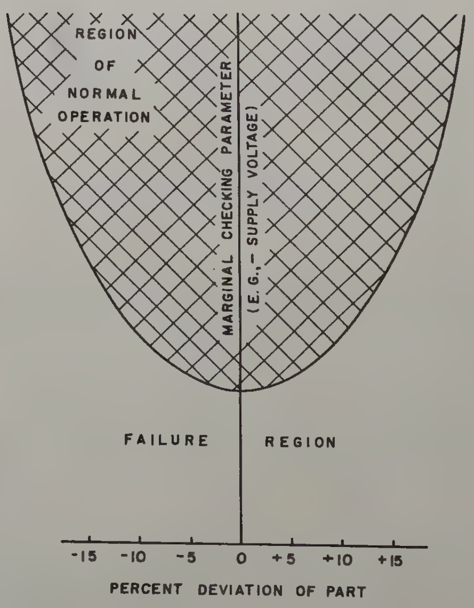


Fig. 5—Marginal checking-locus of failure points.

involved in the integration of these parts to form reliable equipment can be enumerated.

The first and most obvious precaution is to keep it simple. Granting that a given equipment will require a minimum degree of sophistication, the fact remains that there is still plenty of opportunity to gild the lily. The temptation is great for our bright, inventive designers to emulate the Rube Goldberg approach; this author having done design work fully sympathizes with them, and realizes that designing for performance is much more interesting and glamorous than designing for reliability. However, the latter is one job that cannot be bypassed.

Equipment should not only be simple in design, but simple to operate. One of the causes of equipment unreliability is the maladjustment of controls because of the excessive number of front-panel adjustments which require an engineer's training to be correctly set. This is due to a design tendency to include controls which, when properly adjusted, increase equipment performance levels somewhat, but when maladjusted, reduce the equipment function to almost inoperable levels.

In addition to operability, the equipment should be designed for a high level of maintainability. The latter is defined as the reciprocal of the mean net time to repair failures. Expressed mathematically, $M = \frac{1}{x}$

$$\text{where } x = \sum_{i=1}^{n} \frac{x_i}{n}.$$

Because of the increasing complexity of equipment and the decreasing quality level of service personnel, it is necessary to make equipment easy to maintain. This presupposes the adoption by the designer of a disposal-at-failure maintenance philosophy. Circuits are designed as modules for ease of troubleshooting and replacement. In accordance with this philosophy are the employment of printed circuitry, encapsulation and miniaturization. It is recommended that specific maintainability requirements be included in development contracts for equipment.

Another important principle is the practice of conservatism of electronic circuit design. Parts should be derated and tube voltages should be selected so that the lowest values which give the required performance will be employed. The latter step improves reliability in several ways. Part failure is minimized because of reduced peak currents, lowered potential stress and decreased heat dissipation. The likelihood of avalanche failure is greatly reduced. In addition, the consequent restricted energy level reduces the incidence of parasitic oscillations. One manufacturer of television receivers who was notorious for designing for stresses well above the reliability limits in order to conserve parts is brought to mind. Not only was the reliability of the receivers extremely low, but the maintainability was so poor that most servicemen were extremely reluctant to work on them. This approach is one certainly not to be recommended for military equipment.

Redundancy is one method employed as a reliability measure. Two or more identical parts are placed in parallel so that failure of one part will not make the equipment inoperative. This is akin to moving the pitcher, shortshop, second baseman and centerfielder all into line to field a ground ball or to the use of both suspenders and a belt to keep one's trousers from falling. Redundancy is a necessary evil and is recommended only for critical parts where every effort to achieve the required part reliability has failed.

Reliability considerations require that the parts in a circuit be integrated into a package which is designed with a view towards optimizing ruggedness and thermal adequacy. With respect to ruggedness, the design should be such as to restrict the maximum vibrational transmissibility (transmissibility is the ratio of induced to applied vibration amplitude) to a value as near to unity as possible. For example, the use of the clamped-clamped type of assembly, where mounting boards are clamped at both ends, rather than the cantilever type, is recommended. The basic principle of adequate thermal design is to make the total equivalent thermal resistances from all heat generating parts to the thermal sink or environment as low as possible. That is, adequate conduction, convection and radiation paths should be provided to dissipate the heat. In most circuits, the electron tubes are the principal heat-generating parts and their operating temperatures generally exceed the permissible operating temperatures of the other parts. Therefore, thermal adequacy begins with tube location. It is desirable to locate the tubes as far as possible from the parts having the lowest permissible operating temperatures. Employment of equivalent thermal circuit diagrams in a paper analysis is of great assistance in minimizing cut-and-try methods in design. Proper packaging for ruggedness and thermal adequacy is a very important step toward equipment reliability.

Employment of standardized electronic circuitry by equipment designers can be very effective in achievement of high equipment reliability. The National Bureau of Standards and the Navy Elec-

tronics Laboratory have designed a variety of electronic circuits with the emphasis on a high order of reliability. These have been published in the "NBS Preferred Circuits Handbook" and the "NEL Reliability Handbook." A recent study of 83 pieces of Navy electronic equipment showed that 30 per cent of the circuitry could be performed by the circuits listed in these handbooks. It is recommended that the development of preferred circuits be extended and that equipment designers get into the habit of using these as much as possible. This may be a blow to the pride of designers who make a fetish of originality, but it will also be a blow struck against unreliability in their equipment. The use of standard assemblies which may be used in many equipments leads to a further gain in reliability because of the improved quality control which accompanies higher production levels.

Proper liaison is an important factor and its omission can contribute to unreliability. Liaison between designer and user is desirable to acquaint the designer with the user's environmental, operating and maintenance problems. This is much more important with military than with commercial equipment since, in the case of military equipment, a specific number of equipments are contracted for and manufactured before there is feedback from the user to the producer informing the latter of equipment shortcomings. However, in the case of commercial equipment, feedback begins with the first shipments of equipment so that design weaknesses can be corrected before large scale production takes place. This author recalls once having to redesign a piece of equipment after pilot production had begun simply because the designer had not been aware of the conditions of operation of the equipment with the result that the latter proved to be unreliable for the intended application. Proper liaison with the user would have obviated the difficulty.

Liaison among the various groups involved in the development of an equipment is important, too. RCA uses an elaborate system to ensure maximum reliability. After the development contract is awarded, the design engineer must justify his ideas before a panel of experts—reliability engineers, parts people, specialists in shock, vibration and heat, circuit designers, etc. This is done before the design is started and also after the breadboard is ready. When the model is constructed, it is thoroughly tested, the results being reviewed by experts and analyzed in terms of the whole system. Weak points and lack of reliability are spotted. Undesirable interaction effects between various components of the equipment are eliminated. If found necessary, other tests are recommended, circuits are modified, packaging is changed. In the end, the equipment functions according to specified requirements. All of this review may seem to be unnecessarily time-consuming, but this procedure produces very large savings in re-engineering costs and what is more important, results in a highly reliable product. Emulation of this philosophy is definitely recommended for all developers of military electronic equipment.

PILOT PRODUCTION

Pilot production follows development of the equipment, and its primary purpose is to enable the customer to get an idea of what may be available from regular production. It is also of benefit to the manufacturer in that it permits him to prove out the tooling and manufacturing processes.

In addition to the usual performance tests, a battery of environmental tests should be carried on to determine the reliability. Among these should be temperature and input voltage variation, vibration and off-on cycling as a minimum. Other environments selected depend on the corresponding service conditions and may include humidity, salt spray, sand, dust, shock, radiation, etc.

Because of the inherent characteristics of the pilot production process, the output is unavoidably heterogeneous and the reliability tests are indicative of the capability of the manufacturing process rather than of acceptability.

MANUFACTURE

We will now consider the area of manufacture, or full production. The most obvious method of assuring that unreliabilities do not creep in during the manufacturing process is by practicing adequate quality control.

Another step which can be as important is to survey and rate the vendors in the field, qualifying their products.

Automation can be of help in improving reliability. Investigations indicate that mechanized assembly techniques for electronic equipment tend to maximize reliability. These techniques include processed wiring circuitry, mechanized insertion of parts, automatic mass soldering and automatic functional testing. Mechanized production and testing methods possess an advantage over manual methods in that the former avoid the irregularities in techniques and materials of the latter, resulting in improved reliability.

One of the reasons for the existence of unreliable equipment is the tendency to rush it into production before the development has really been completed. Present procurement practices,

which aim to provide accelerated delivery of electronic equipment, tend to minimize the time allowed for adequate reliability evaluation. This telescoping of development with procurement is accomplished at the expense of a sound reliability test program during the vital engineering phase and must necessarily be reflected in decreased reliability of the end product. Therefore, it is recommended that production be postponed until adequate engineering tests prove that the item in question fully meets the reliability requirements.

TRANSPORTATION

The transportation phase furnishes an excellent opportunity for introduction of unreliabilities. The military services have experienced substantial damage to equipment during shipping, resulting from improper packaging and packing. Since the damage which occurs is not always detectable and therefore repairable, incipient failures may easily occur. Proper packaging and packing is an important link in the reliability chain.

The steps taken to insure proper packaging design of the equipment to withstand operating shock and vibration will also serve to protect it during transportation. In addition, it is necessary to investigate the shock and vibration experienced by equipment packed and shipped in containers. It is recommended that instruments be developed which will record the amplitudes and durations of shocks to which equipments are subjected in shipment. These should be of a type which will operate unattended for a period of several weeks.

In addition, it will be necessary to determine specific cynamic values for a wide variety of cushioning materials for the use of designers of shipping containers.

The recommended research in packaging and packing should lead to increased operational reliability of electronic equipment.

STORAGE

Since production contracts provide for sufficient numbers of equipments not only to meet the current operational requirements but also to allow an adequate reserve, it is obvious that the excess must be kept in storage for appreciable periods of time. This process subjects these items to the deleterious effects of corrosion, chemical action and other forms of deterioration, thus posing an additional reliability problem.

Equipment should be stored under conditions which minimize rate of deterioration. This sounds very simple, but it is a fact that these conditions can only be made known by: 1) accelerated aging tests, and 2) monitoring of items in storage. Accelerated aging tests are necessary to obtain data in a relatively short time during equipment development. However, since the conditions encountered during storage cannot be perfectly simulated, they are not a completely satisfactory substitute for storage monitoring.

In order to provide data on deterioration in a form which is readily usable by: 1) the agency directly concerned with the given item, and 2) agencies which require such data as background information for similar items or subdivisions thereof, it is essential that such data be made available in convenient form. The most suitable forms are considered to be punch cards or magnetic tape. At present, huge masses of information are buried in miscellaneous and heterogeneous reports in the archives of multitudes of agencies. As the number of equipments in existence increases, this situation will become greatly aggravated unless a streamlined system of data reporting and reduction is adopted. It is recommended that a working group be established at Department of Defense level to develop such a system of data handling which will be uniformly employed by the various services and will be designed to be compatible with the requirements of all the agencies. However, if such a plan is to succeed, it must be implemented by directive at DOD level which will make use of the adopted system mandatory.

In order to provide maximum benefits from such a system, it would be advisable to retain a life history of each individual equipment from the time it is manufactured until it is removed from service by the operating unit. Only in this manner can a comprehensive knowledge of the variation in condition be obtained. The information that is obtained by the reporting sources will be transcribed to forms suitable for handling by computing machines and will be subjected to statistical and engineering analyses. Results obtained will be in a form that can be used directly by designers, manufacturers, storage personnel or operating agencies. The Ballistic Research Laboratories are presently working on such a system for use with military electronic equipment. Collaboration with groups working on the same problem is invited.

In order that the data obtained be valid, it is essential that the equipment used to test these items be of sufficient precision. Although the specifications for equipment in storage and operation are generally less stringent than those for acceptance, this should not imply that a corresponding decrease in precision of test sets used at these stages is permissible; rather, all test sets used for any given equipment should be of similar ac-

curacy, regardless of whether employed in the acceptance, storage or operational phases. Only in this manner can trustworthy and comparable data be obtained.

Accuracy of measuring equipment presumes calibration against precise standards. In the interests of obtaining uniform results, it might be desirable to appoint a panel at DOD level to prescribe calibrating equipment to be employed. In fact, it may even be advisable to institute a Military Bureau of Standards, similar in function to the Calibration Division of the National Bureau of Standards but slanted towards calibration of the type of test equipment employed by the Armed Forces.

Another prerequisite for assuring validity of data is employment of high caliber technicians in the organizations performing the reporting function. This is contingent upon acceptance of the recommendations included in the Cordiner report dealing with the shortage of trained technicians in the Armed Forces. More will be said about this problem in the portion of the paper devoted to maintenance.

Another cause of insufficient reliability is the fact that the design of suitable test equipment is usually treated as a secondary consideration. The testers are often not available until it is much too late to be of use in assuring reliability of the item. It is necessary that design of the basic equipment and its testers be treated integrally.

OPERATION AND MAINTENANCE

Maintenance is an important factor in the effort to achieve reliability, its purpose being to sustain designed performance and continued operation of equipment and systems in order to attain the highest degree of operational readiness.

Maintenance of electronic equipment is dependent upon such factors as equipment maintainability, personnel training, preventive maintenance procedures and quality of support material such as technical manuals, test equipment and test facilities.

Maintainability has already been defined in this paper as the reciprocal of the mean net time to repair failures. Unfortunately, there is nothing that maintenance personnel can do about this characteristic, since it is predetermined. If the design people have been careful to observe the tenets of the disposal-at-failure maintenance philosophy such as modular construction, encapsulation, etc., then the maintainability of the equipment should be high.

Even if the design of the equipment enables relatively unskilled personnel to perform the maintenance function at the lower echelons, highly skilled technicians are still needed at the top echelons. Unwise policies have permitted the situation to deteriorate to the point where large numbers of extremely expensive equipments are at the mercy of fewer and less skill personnel than ever before. This grave situation can be alleviated only by taking immediate and drastic steps. The most effective one would be the decreasing of the high turnover rate of trained men by offering sufficient incentive to remain in the services. This could be accomplished by raising the pay scales to realistic levels and by reinstituting the many fringe benefits which once were enjoyed. To assure that ability rather than longevity should be the basis for promotion, a merit system should be adopted. Another very effective means of maximizing available skilled manpower is the elimination of the practice of requiring the technician to perform nontechnical routine duties, such as K.P., guard duty, etc. A less direct, but nevertheless important factor is the low level of technical background possessed by the average recruit, necessitating inordinately long training periods acquiring basic knowledge which should have been obtained previously. It is therefore advantageous to the Department of Defense to seek the adoption of better and more thorough training in mathematics and the physical sciences at the secondary school level.

Reliability can be greatly increased by detecting potential failures before they have an opportunity to occur. One of these maintenance techniques is called marginal checking and is related to the marginal checking performed during design.

The principle underlying marginal checking of electronic equipment as a preventive maintenance procedure is as follows. If all the parts are in good condition, then variation of parameters, generally power supply or signal voltages, will not cause the equipment to fail. However, failure may be induced if a part has deteriorated; *e.g.*, if the transconductance has been appreciably reduced. The method employed is to vary voltages between specified limits and observe whether the equipment functions properly. For instance, in the checking of a computer, a problem may be fed to it while varying the voltage on portions of the computer in turn. An incorrect answer serves to localize the maloperating circuit and then the potentially defective part.

Another technique which is being investigated is the prediction of imminent failure based on the variation of parameters of certain electronic parts

Armour Research Foundation, under Air Force contract, is conducting studies which indicate that resistor noise progressively increases prior to failure. It also appears that a decrease in insulation resistance of both resistors and inductors may be a harbinger of failure. Thus, monitoring of the equipment may provide a means of preventing failures by furnishing sufficient warning to permit part replacement. It is suggested that research along these lines be expanded, since application of these principles will be of great assistance in improving reliability.

Maintenance of electronic equipment requires the use of precision test equipment for a variety of measurements. Calibration of testers must be dependable regardless of time or location. This requirement imposes a need for uniform calibration standards throughout all military installations. It is recommended that DOD set up calibration centers in selected areas. Utilizing the National Bureau of Standards for primary reference, these centers would service all military agencies within their respective areas.

Support material such as test equipment, training, and instruction manuals are essential for the proper performance of the maintenance function. Yet, more often than not, these are not available simultaneously with the main equipment. Operation of the latter without the guidance furnished by the applicable technical manuals and use of the proper test equipment is not conducive to achievement of maximum reliability. Therefore, it is recommended that no equipment be released for distribution unless accompanied by the applicable support material.

The scope of this field is tremendous, so that only the highlights have been touched upon in this paper. However, if the recommendations which have been made were universally adopted, this author believes that the reliability of our military electronic equipment would be greatly increased. In fact, if this accomplishment were made known to the enemy, it might even serve as an effective deterrent to military conflict.

REFERENCES

[1] Advisory Group on Reliability of Electronic Equipment (AGREE), Office of Assistant Secretary of Defense (OASD), "Reliability of Military Electronic Equipment," Rept.; June 4, 1957.

[2] R. Lusser, "The Notorious Unreliability of Complex Equipment," Redstone Arsenal, Huntsville, Ala., Rept.; September, 1956.

[3] R. Lusser, "Reliability of Guided Missiles," Redstone Arsenal, Huntsville, Ala., Rept.; September, 1954.

[4] S. G. Bassler, "Principles of electronic circuit packaging," *Electrical Manufacturing;* August, 1955.

[5] S. G. Bassler, "Electronic circuit packaging for missile applications," *Electrical Manufacturing;* March, 1958.

[6] R. Lusser, "Which road to reliability?" *Electronic Equipment;* January, 1957.

[7] R. Lusser, "A Study of Methods for Achieving Reliability of Guided Missiles," U. S. Naval Air Missile Test Center, Point Mugu, Calif. Tech. Rept. No. 75; July 10, 1950.

[8] R. Lusser, "General Specifications for the Safety Margins Required for Guided Missile Components," U. S. Naval Air Missile Test Center, Point Mugu, Calif. Tech. Rept. No. 84; July 10, 1951.

[9] R. Lusser, "Planning and Conducting Reliability Test Programs for Guided Missiles," U.S. Naval Air Missile Test Center, Point Mugu, Calif. Tech. Rept. No. 70; June 20, 1952.

[10] M. A. Acheson, "The unreliable universal component," *Electronic Equipment;* January, 1957.

[11] K. A. Pullen, "A new approach to conservative design," *Electronic Equipment;* May, 1957.

[12] C. J. Savant, and H. S. Hansen, "Reliability as a responsibility of engineering management," IRE TRANS. ON RELIABILITY AND QUALITY CONTROL, No. PGROC-9, pp. 45-48; January, 1957.

[13] R. C. Marder, "The effect of mechanized production techniques upon reliability," *Military Automation;* January-February, 1958.

[14] Department of Defense Instruction Number 3232.2. "Electronic Equipment Failure Data Reporting System and DD Forms 787 and 787-1," Rept.; February 23, 1956.

WHAT PRICE UNRELIABILITY

DANA A. GRIFFIN†

PART I

The continuing record of mission failures of our complex ICBM and IRBM missiles is highlighted by every communication medium in the world. Similar mission failures of other types of complex systems are not publicized. We can be sure that they are extremely costly because there are many more of them.

The excessive costs of such unreliability might be written off as a necessary means to the end, if this were the only factor to be considered. Unfortunately, this is not the case. Mission success in the ICBM-IRBM missile and the antimissile field depends upon the ability to act instantaneously without failure.

The purpose of this article is to demonstrate the need for a realistic look at the mission success potential of our complex weapons systems and the way to obtain a large increase in mission success potential at much lower costs.

Current military procurement practice is to establish a reliability requirement on every weapon system and then assume that the prime contractors will attain the desired level of performance using presently available parts. This policy establishes a defense posture based on wishful thinking rather than reality insofar as our complex systems are concerned. The basic building blocks of all systems, the small component parts, do not possess the life expectancy to meet the system reliability requirements.

The moment we express permissible system failures in terms of specific missions, we must automatically consider the life expectancy of every part of the system that is likely to fail and by so doing, cause a mission failure.

Computations on many systems now in development or production clearly indicate that the reliability requirements cannot be met with a high degree of confidence. The failure rate of the majority of the electronic component parts presently available for use in these systems is from 10 to 100 times poorer than the desired system failure rate will allow.

Let us distress the statisticians with three oversimplified expressions of an elementary formula:

†The Daven Company, Livingston, N. J.

$$\frac{A}{B} = C, \quad A = B \times C \quad \text{or} \quad \frac{A}{C} = B$$

If A equals the permissible mean time between failure rate of the system in hours, and B equals the number of parts whose individual failures can cause a mission failure, C can be expressed as the permissible number of part failures in a given number of hours.

If we know the value of A and the number of parts in the system B, we can determine the value of C. Similarly, if we know the number of parts involved, B, and their failure rate, C, we can determine A, the system rate. If we know A and C, we can determine the maximum number of components we can employ in a system to obtain the requisite system failure rate.

For the past two years, contracts have been let for complex weapon systems with values assigned to A that are required to assure mission success, without regard to the part count, B, or component part failure rates, C.

With a simpler mathematical table (Table I) we can get to the nub of the situation. Component part failure rates are expressed in terms of their Acceptable Reliability Levels (ARL) which permit a certain number of failures per thousand hours of operation in a given quantity of parts. These are expressed in ARL percentages as indicated below.

TABLE I

ARL Percentage	Number of Failures per 1000 Hours per 100,000 Parts
1	1000
0.1	100
0.01	10
0.001	1

If a missile system employes 100,000 parts and the permissive MTBF rate for the system is 100 hours, we need parts with a failure rate of 0.01 per cent to obtain the specified system failure rate. Electronic components with these failure rates may be available in the Soviet Union, but they are conspicuous by their absence in the United States.

The basic reason for this unfortunate situation

is the failure of our military procurement personnel to recognize the nature of our free enterprise system. There are two independent tiers of manufacturers, broadly speaking. The first tier produces systems and the second tier produces many of the critical component parts that make up these systems.

Billions of dollars have been allocated to first-tier contractors to produce systems using presently available, relatively unreliable component parts. A companion program of financial aid to second-tier component part manufacturers for component part improvement to the levels required by first-tier contractors does not exist.

The failure to provide for a component part improvement program has created an unprecedented situation. The manufacturers of a number of complex systems require component parts that meet a 0.01 per cent ARL. They cannot buy such parts at any price from component part manufacturers at the present time.

Bankruptcy via Redundancy

At first glance, it might be assumed that enough unreliable missiles can be purchased to offset their lack of adequate MTBF rates. The hopelessness of this approach can be illustrated by a purely hypothetical example.

An effective antimissile, missile shield for our continent might require 1000 missiles ready to fire. Their complexity might demand the use of component parts with an ARL of 0.001 per cent in order to provide the necessary mission success potential.

In order to approach this goal with missiles using parts with a 0.1 per cent ARL, 100,000 missiles will be needed! Ignoring the cost of the extra ground installations and the full-sized army to repair and operate the unreliable missiles, an approximation of production costs may prove of interest as it points the way to the justification of large-scale program for component part improvement.

As an example, we might buy 100,000 missiles (ARL 0.10 per cent) at $1,000,000 each for $100,000,000,000. Or, by arbitrarily increasing the unit cost ten times to insure reliability, we might buy 1000 missiles (ARL 0.001 per cent) at $10,000,000 each for $10,000,000,000. We must spend an additional 90 billion dollars for enough unreliable missiles to approach the same mission success potential, and we must ignore the time wasted in firing 99 abortive shots out of 100 which will increase the probability of enemy breakthrough by a substantial amount.

There is no reason to assume that it will cost 9 billion dollars to raise the ARL from 0.1 per cent to 0.001 per cent. However, it will require substantial expenditures at the second-tier level of component parts manufacturers for development and facilities. This is by no means the only justification for a major attack on the problem of improving component part life expectancy to levels undreamed of a few years ago.

Maintenance Costs

Major improvements in component part life expectancy can reduce our annual maintenance and repair bills by many billions of dollars.

The high cost of replacing short-life component parts with similar short-life parts can be demonstrated by elementary multiplication and the failure rates listed in Table II.

As we produce systems requiring billions of parts every year, plus billions more for replacement purposes, replacement costs per billion parts will be estimated for the various ARL percentages.

TABLE II

ARL percentage per 1000 Hours	No. of Failures per Billion Parts	Unit Replacement Cost	Cost per 1000 Hours	Cost for 50,000 Hours (System Life)
1	10,000,000	$20.00	$200,000,000	$10,000,000,000
0.1	1,000,000	20.00	20,000,000	1,000,000,000
0.01	100,000	20.00	2,000,000	100,000,000
0.001	10,000	20.00	200,000	10,000,000

If we assume a current failure rate at the 0.10 per cent level, a change to the 0.01 per cent level will save $900,000,000 in repair costs per billion parts in a 5-year period. We have reason to believe that the unit replacement cost will be challenged. If all cost factors are considered, probably it should be increased by a substantial amount. The dollar-wise accuracy of the calculations are not particularly important, however. We are only considering 1 billion parts in a military hardware system that employs and stocks many billions of parts.

The important factor is the huge reduction in repair costs that will obtain by decreases in failure rates anywhere in the range between the 1 per cent ARL and the 0.01 per cent ARL.

Realistic figures on weapon system mission success probability can only be obtained with a knowledge of the ARL percentages on all categories of parts going into the systems, plus a count on the number of each category of part.

A continuous survey of ARL percentages currently available in all part categories will give the Defense Department a powerful measuring tool that can be used to determine realistic system re-

liability levels and also serve as the basis for a large-scale program for the improvement of component part life expectancy that is so sorely needed.

PART II

In Part I of this paper, the need for a major improvement in the life expectancy of the parts going into our weapons systems was demonstrated, and the justification for major expenditures in this field was developed in two areas. Many billions of dollars can be saved by a reduction in the number of systems required to insure mission success, and system repair costs can be diminished by a substantial increase in component part life expectancy.

In Part I it was suggested that our weapons systems reliability be reappraised from the bottom up. That is, to use the life expectancy levels of presently available component parts in system reliability predictions. The difference between the desired system mean time between failure rates, specified by our tactical experts in our weapon system contracts, and the realistic rates which such computations will disclose, should give some cause for alarm to those charged with the responsibility of protecting our major cities from total obliteration and winning a major war should the occasion arise.

Unfortunately, we cannot buy time, so in order to dispel possible complacency on the part of our tactical commands, it is necessary to point out that the immediate implementation of a component part improvement program cannot possibly result in improved systems for 3 to 5 years.

The reasons for this time delay will be discussed later. It is important that the tactical commands realize that they are being forced to use complex systems now in production or in the development stage which will not provide the specified mission success potential and that no substantial improvement can be expected for some time to come!

We have the "know-how" and the facilities to build complex systems, but we don't have the "know-how" or the facilities to build the reliable parts that these complex systems require in order to function in the specified manner.

There are many reasons for this deplorable state of unpreparedness. The major factors are:

1) The state of art in component part development.
2) Failure to recognize the scope of the problem.
3) The lack of adequate funds and of willingness to spend them at the proper level in our two-tier industry.

It would be easy to write a full-length volume on the history of electronic component part development for the past forty years. In brief, there have been two motivating factors: the needs of the home radio-television industry and the military services. The volume production demands of the former are responsible for the capital investment for production facilities and much of today's "know-how." Military expenditures have been almost exclusively confined to the areas of basic research and the development of parts that would work under field conditions.

These combined efforts have enabled us to reach a plateau where the ARL's of most parts range between 0.5 per cent and 0.1 per cent. This is more than adequate for the needs of the radio and television industries. Further improvement in component part life expectancy must be wholly financed by the military services.

A shift from these ARL's to a new plateau of a 0.01 per cent ARL for all categories of component parts is an innocent-looking expression. Actually, its attainment will require major technological breakthroughs in physics, chemistry, product design, and process control, to name a few areas where current knowledge and techniques are inadequate.

Today, the services are buying electronic complexity in varying degrees in modern weapon systems; production can no longer be evaluated in terms of dollars per pound of product. These production yardsticks of World War II are completely inadequate.

The importance of increased component part life expectancy in complex systems can be illustrated with Table III which is completely hypothetical.

TABLE III

Name	Number of Parts Approximately	ARL Percentage Required	ARL Percentage Available	Mission Success Potential
Titan	100,000	0.01	0.1	Very Poor
Atlas	50,000	0.02	0.1	Poor
Thor	10,000	0.1	0.1	Good
Jupiter	15,000	0.05	0.1	Fair
Polaris	75,000	0.01	0.1	Very Poor

The substitution of real numbers in a table of this type on all of our complex weapons systems would be enlightening, to say the least. The mission success potential is evaluated in terms of the disparity between an optimistic value of a 0.1 per cent ARL for available component parts and the

component part ARL required to meet the specified MTBF rate for the system in question.

Component Part Survey

As suggested in Part I, the first sensible corrective steps will be to survey available ARL's for every category of part and obtain a count on the number of parts of each category going into each system. From these data, grand totals on each category of parts required by the services can be obtained and the scope of the problem can be established.

Facilities Funding

From this survey we may find, for example, that 50,000,000 transistors with an ARL of 0.01 per cent are needed between 1961 and 1962. Since fully automatic production facilities are required to meet the ARL of 0.01 per cent, funds for this purpose must be provided just as machine tools and billions of dollars' worth of other facilities are supplied to first-tier prime contractors.

The question of how many contractors and which of the many potential contractors will be given facilities must be resolved, plus hundreds of other questions in many areas. An effective, well-coordinated program covering all categories of component parts will require an efficient administrative organization within the Defense Department at the decision-making level, which presently does not exist.

There are a number of other factors that affect weapons systems performance adversely. Errors in design, factory malpractice, inadequate inspection, improper installation, and poor maintenance play their part in the reduction of mean time between failure rates. Any of these faults can be corrected in a relatively short space of time. This is not the case insofar as major improvement in component part life expectancy is concerned. Basic research, part development and the design and fabrication of the requisite production facilities will take large amounts of time even though a so-called crash program is instituted to accomplish the desired results.

CONCLUSIONS

1) Our defense posture is impaired by the degradation of weapon system mean time between failure rates below the desired levels. This is occasioned by the lack of component parts with adequate life expectancy.
2) We are unprepared to produce component parts with the life expectancy required.
3) It will take from 3 to 5 years to obtain a major improvement in component part life expectancy after the initiation of a large-scale program in this area.
4) A survey of available component part acceptable reliability levels can be used to determine realistic weapon system failure rates and the scope of task of improving component part life expectancy to the levels required by our complex weapons systems.
5) Corrective action by the Department of Defense is essential. It will enhance our defense posture and save billions of dollars annually in production and maintenance costs.

CRITERIA FOR DETERMINING OPTIMUM REDUNDANCY*

R. E. BARLOW† *and* *L. C. HUNTER*†

Summary—Redundant circuits whose components may suffer either an open-circuit or a short-circuit type of failure are considered. A probabilistic model for such circuits is proposed. Two criteria for determining optimum redundancy are studied. A formula for obtaining the number of components which maximize reliability is derived for general failure distributions. A table for obtaining the number of components which maximize the expected life of the circuit is presented for the case of exponential failure.

*This work was prepared for the U. S. Army Signal Research and Development Lab. under Signal Corps Contract No. DA 36-039 SC-78281.

†Electronic Defense Lab., Sylvania Electric Products, Inc., Mountain View, Calif.

INTRODUCTION

By a redundant circuit, we shall mean any circuit all of whose primary components perform the

same function. For example, an arrangement of n switches or diodes could constitute a redundant circuit. We shall be mainly concerned, however, with parallel and series circuits and arrangements built up of such units. Continuing the discussion from a previous report [1], we shall denote the reliability of the circuit at time t by $R(t)$. This is understood to be the probability that the circuit is operating at time t, given that it was put into operation at $t = 0$. We agree to count time only while the circuit is operating. Since we are not concerned with repair, the failure distribution of the circuit is given by

$$G(t) = 1 - R(t).$$

Knowing G, the expected life of the circuit can be computed.

It is usually assumed that increased redundancy assures increased reliability. However, in practice, components in parallel or series often fail in such a way as to effect the entire circuit. For example, given a network of diodes in parallel, one could short-circuit and render the whole network inoperative even though one or more of the remaining diodes were in operating condition.

We shall be concerned with redundant circuits whose components can fail in either of the following two ways:

1) An open-circuit failure can occur;
2) A short-circuit failure can occur.

Good examples of the first type of failure are: a diode may fail and not allow current to pass in either direction, or an open switch may fail to close when desired. Examples of the second type of failure are easy to find. A diode may fail and allow current to pass in both directions, or a closed switch may fail to open when desired.

Lipp [3], Price [5], and others have considered probabilistic models which contain the features described above with the exception that failure distribution functions were not explicitly considered. Our model also differs in other important respects. Moskowitz [4] has given an adequate analysis of redundancy networks but does not consider components which can fail in either of two ways.

In order that our calculations be valid, we shall need to make the assumption that circuit components are independent in a probability sense. This precludes the application of our results to many important circuits. Gilmore and Levi [2] have considered the problem of component independence in terms of adequate isolation. They indicate that such isolation is practical in a vacuum tube design but that transistors are not as easily isolated. They point out that redundancy techniques may even be detrimental rather than beneficial in some cases.

MAXIMIZING RELIABILITY

Let F_i denote the failure distribution of the ith component in a redundant circuit where F does not distinguish between types of failure. Let p_i denote the probability of an open-circuit failure in the ith component given that it has failed. We note that p_i is a conditional probability. Similarly, we let $1 - p_i$ denote the probability of a short-circuit failure in the ith component given that it has failed. Let U denote a unit consisting of m components in parallel and let V denote a unit consisting of n components in series.

Suppose all components of U have identical failure distributions. Then it is easily seen that, for a parallel circuit

$$\begin{aligned} R_p(t) &= \sum_{R=0}^{m-1} \binom{m}{k} [1 - F(t)]^{m-k} [pF(t)]^k \\ &= [1 - F(t) + pF(t)]^m - [pF(t)]^m \qquad (1) \\ &= x^m - y^m . \end{aligned}$$

Note that $x^m - y^m$ is non-negative and zero for $m = 0$ and $m = \infty$. Since the derivative of this expression with respect to m set equal to zero has a unique solution, it must be a maximum. Hence, the optimum integer m is close to the value

$$m^* = \log[(\log y)/\log x]/\log(x/y).$$

In general, $m = 1$ is best whenever $p \leqslant \frac{1}{2}$, as one would expect. To see this, note that $p \leqslant \frac{1}{2}$ implies

$$F(t) \geqslant 2pF(t)$$

or

$$1 \geqslant 1 - F(t) + 2pF(t) = x + y$$

and

$$x - y \geqslant x^2 - y^2 .$$

Since the expression $x^m - y^m$ is continuous in m and has a unique maximum when we allow m to assume all positive real values, the result follows.

Let all components of V have identical failure distributions. Then, substituting $(1 - p)$ for p in (1), we have, for a series circuit,

$$\begin{aligned} R_s(t) &= [1 - pF(t)]^n - [(1 - p)F(t)]^n \\ &= w^n - z^n . \qquad (2) \end{aligned}$$

The optimum n for fixed p and $F(t)$ is the same as before when we substitute w for x and z for y. In general $n = 1$ is best whenever $p \geqslant \frac{1}{2}$.

Proceeding in the same way, we can determine the reliability of more complicated networks. For convenience we make the following definitions.

Definition 1

A parallel-series (PS) arrangement shall denote m type V units in parallel.

Definition 2

A series-parallel (SP) arrangement shall denote n type U units in series.

Consider m type V units in parallel. We wish to determine the reliability of a parallel-series arrangement. To do this, let

a = probability a unit has not failed,
b = probability a unit has failed favorably,
c = probability a unit has failed unfavorably.

In this case, V fails unfavorably if all n components in V short-circuit. Again, let all components have identical failure distributions. Then

$$\text{(PS)}: \quad R_{ps}(t) = [a + b]^m - b^m .$$

Since $a + b + c = 1$, and

$$c = [(1 - p)F(t)]^n$$

$$a = [(1 - F(t) + (1 - p)F(t)]^n - [(1 - p)F(t)]^n$$

we have

$$\text{(PS)}: \; R_{ps}(t) = [1 - \{(1 - p)F(t)\}^n]^m - [1 - \{1 - pF(t)\}^n]^m .$$

A similar argument showing the reliability of a series-parallel arrangement is

$$\text{(SP)}: \; R_{sp}(t) = [1 - \{pF(t)\}^m]^n - [1 - \{(1 - p)F(t)\}^m]^n .$$

Both expressions can be optimized over m and n by setting the partials, with respect to m and n respectively, equal to zero and then solving. The optimum values will, of course, depend on $F(t)$.

If $p = \frac{1}{2}$, we can obtain optimum reliability with either a (PS) or (SP) arrangement. To see this, fix $F(t)$ and let $R_{ps}(t) = f(m,n)$ and $\max_{m,n} f(m,n) = f(m^o,n^o)$. Let $R_{sp}(t) = g(m,n)$ and $\max g(m,n) = g(m',n')$. Note that for $p = \frac{1}{2}$, $f(m,n) = g(n,m)$. Since

$$f(m^o,n^o) \geqslant f(n',m') = g(m',n')$$

and

$$g(m',n') \geqslant g(n^o,m^o) = f(m^o,n^o)$$

we have $f(m^o,n^o) = g(m',n')$.

Hence, $m^o = n'$ and $n^o = m'$. In this case we can obtain optimum reliability with either a (PS) or (SP) arrangement.

Suppose now that the components do not have identical failure distributions, but that they still perform the same function. Then, reasoning as before, we obtain

$$R_{ps}(t) = \left[1 - \prod_{i=1}^{n} (1 - p_i)F_i(t)\right]^m - \left[1 - \prod_{i=1}^{n} \{1 - p_iF_i(t)\}\right]^m$$

and

$$R_{sp}(t) = \left[1 - \prod_{i=1}^{m} p_iF_i(t)\right]^n - \left[1 - \prod_{i=1}^{m} \{1 - (1 - p_i)F_i(t)\}\right]^n .$$

MAXIMIZING EXPECTED LIFE

We have obtained the reliability of parallel and series circuits based on our probabilistic model. It is an easy matter to optimize the reliability function for a given value of t. However, in many cases the desired operating life is indeterminate. Perhaps a better criterion in this circumstance is to optimize the expected time to failure over m or n. If our failure distribution is exponential, then it is an easy matter to calculate this quantity.

If $G_m(t) = 1 - R_p(t)$, then $\int_0^\infty t\,dG_m(t)$ is the expected time to failure for a parallel circuit. Similarly, let $G_n(t) = 1 - R_s(t)$. If $F(t) = 1 - \exp(-\lambda t)$, then it can be shown that

$$\int_0^\infty t\,dG_m(t) = 1/\lambda \sum_{k=1}^{m} \frac{p^{m-k}}{k} \tag{3}$$

for a parallel circuit and

$$\int_0^\infty t\,dG_n(t) = 1/\lambda \sum_{k=1}^{n} \frac{(1-p)^{n-k}}{k} \tag{4}$$

for a series circuit. This pleasing result leads to the observation that the optimal m or n can be determined solely on the basis of a knowledge of p. Let p_m be the solution in p to

$$1/\lambda \sum_{k=1}^{m+1} \frac{p^{m+1-k}}{k} - 1/\lambda \sum_{k=1}^{m} \frac{p^{m-k}}{k} = 0. \tag{5}$$

The solutions p_m determine critical intervals for p. Table I has been constructed for values of m and n up to 10. Of course, one should use a series circuit for values of p less than 0.5 and a parallel circuit for values of p greater than 0.5.

As an example of how an engineer might use

this table, suppose that he desired to parallel a number of switches in order to increase the overall reliability. Furthermore, suppose that of those switches that had failed in the past, 95 per cent had been open-circuit failures. Assuming that failures occur by chance rather than through wearout, we can examine Table I to determine the optimal number of switches. From the table it can be seen that $m = 9$.

TABLE I

PARALLEL AND SERIES CIRCUIT PARAMETERS WHICH MAXIMIZE EXPECTED LIFE

Optimal m for a Parallel Circuit	Optimal n for a Series Circuit	Critical Intervals for p
1	more than 10	$p < 0.040$
1	10	$0.040 \leq p \leq 0.046$
1	9	$0.046 \leq p < 0.054$
1	8	$0.054 \leq p < 0.063$
1	7	$0.063 \leq p < 0.077$
1	6	$0.077 \leq p < 0.096$
1	5	$0.096 \leq p < 0.125$
1	4	$0.125 \leq p < 0.175$
1	3	$0.175 \leq p < 0.272$
1	2	$0.272 \leq p < 0.5$
1	1	$p = 0.5$
2	1	$0.5 < p \leq 0.728$
3	1	$0.728 < p \leq 0.8250$
4	1	$0.8250 < p \leq 0.875$
5	1	$0.875 < p \leq 0.904$
6	1	$0.904 < p \leq 0.923$
7	1	$0.923 < p \leq 0.937$
8	1	$0.937 < p \leq 0.946$
9	1	$0.946 < p \leq 0.954$
10	1	$0.954 < p \leq 0.960$
more than 10	1	$p > 0.960$

We now wish to prove (3). Letting $F(t) = 1 - \exp(-\lambda t)$, we obtain

$$R_p(t) = \sum_{k=0}^{m-1} \binom{m}{k} \exp[-\lambda(m-k)t]\; p^k\, [1-\exp(-\lambda t)]^k.$$

Expanding and integrating, we obtain

$$\int_0^\infty t\, dG_m(t) = (m/\lambda) \sum_{j=0}^{m-1} [p^{m-1-j}(1-p)^{j+1} + (-1)^j p^m] \frac{\binom{m-1}{j}}{(j+1)^2}.$$

Note that if

$$\int_0^\infty t\, dG_m(t) = (1/\lambda) \sum_{k=1}^{m} \frac{p^{m-k}}{k}$$

then

$$\int_0^\infty t\, dG_{m+1}(t) = (p/\lambda) \sum_{k=1}^{m} \frac{p^{m-k}}{k} + (1/\lambda)/(m+1).$$

Let $R_m = m \sum_{j=o}^{m-1} [p^{m-1-j}(1-p)^{j+1} + (-1)^j p^m] \dfrac{\binom{m-1}{j}}{(j+1)^2}$.

We assert that $R_{m+1} = pR_m + 1/(m+1)$. If we can show this, the result will follow by mathematical induction. It is a straightforward calculation to see that

$$R_{m+1} = pR_m = \sum_{j=o}^{m} [p^{m-j}(1-p)^{m+1} + p^{m+1}(-1)^j] \frac{\binom{m}{j}}{(j+1)}.$$

Let $k = j + 1$. Then we wish to show that

$$\sum_{k=1}^{m+1} [p^{m+1-k}(1-p)^k + p^{m+1}(-1)^{k-1}] \frac{\binom{m}{k-1}}{k} = \frac{1}{m+1}.$$

Note that $\binom{m+1}{k} = [(m+1)/k]\binom{m}{k-1}$. Hence we need only show

$$\sum_{k=1}^{m+1} \left[\frac{(1-p)^k}{(p)} + (-1)^{k-1}\right] \binom{m+1}{k} = \frac{1}{p^{m+1}}.$$

This can be verified directly.

We now wish to justify the construction of our table. Let

$$S_m(p) = \sum_{k=1}^{m} \frac{p^{m-k}}{k} \quad \text{and} \quad D_m(p) = S_{m+1}(p) - S_m(p).$$

Recall that p_m is the solution in p to $D_m(p) = 0$ [see (5)]. Solutions always exist since $D_m(0) = -1/m(m+1)$ and $D_m(1) = 1/m+1$. It is an easy matter to see that $p_1 = 0.5$ and $p_2 = 0.728$. Also, $D_1(p) = p - \frac{1}{2}$ is clearly increasing in p for $p \geq 0.5$. Hence, $m = 2$ is to be preferred to $m = 1$ in the interval $0.5 < p \leq 1$. It will be sufficient in general to show that $D_m(p)$ is an increasing function of p for $p \geq p_m$. Suppose this assertion is true for $m = k$. Then

$$D_k + 1(p) = p\,D_k(p) - 1/(m+1)(m+2)$$

mplies

$$\frac{d}{dp} D_{k+1}(p) = D_k(p) + \frac{pd}{dp} D_k(p) \geqslant 0 \text{ for } p \geqslant p_k ,$$

since $D_k(p_k) = 0$ and $D_k(p)$ is increasing for $p \geqslant p_k$ by assumption. Since $D_{k+1}(p_k) < 0$, surely $D_{k+1}(p)$ is increasing for $p \geqslant p_{k+1}$. Appealing to the axiom of mathematical induction, we conclude that our assertion is true for all positive integers.

Eq. (4) and the values in the table for series circuits hold by symmetry.

References

[1] R. Barlow and L. C. Hunter, "System efficiency and reliability," 1959 IRE NATIONAL CONVENTION RECORD, pt. 6, pp. 104-110. March, 1959.

[2] J. P. Gilmore and V. H. Levi, "Reliability Through Redundancy," Symposium on Military Electronic Reliability and Maintainability, Rome Air Dev. Center, Rome, N. Y., vol. 3, ASTIA Doc. No. AD-148953, November, 1958.

[3] J. P. Lipp, "Topology of switching elements vs reliability," IRE TRANS. ON RELIABILITY AND QUALITY CONTROL, no. PGRQC-10, pp. 21-34, June, 1957.

[4] F. Moskowitz, "An Analysis of Redundancy Networks," Rome Air Dev. Center, Rome, N. Y., Tech. Note, ASTIA Doc. No. AD-148588; February, 1958.

[5] H. W. Price, "Reliability of Parallel Electronic Components," Symposium on Military Electronic Reliability and Maintainability, Rome Air Dev. Center, Rome, N. Y., vol. 3, ASTIA Doc. No. AD-148953; November, 1958.

INTERVAL ESTIMATION OF PRODUCT RELIABILITY BY USE OF THE NONCENTRAL t DISTRIBUTION

*R. E. SCHAFER**

INTRODUCTION

The use of certain exponential distributions in describing chance failure functions is well known in reliability literature. Methods are available for obtaining "best" estimates of the parameters of these distributions and confidence intervals for the true values.

In many cases, notably in wear-out failure and stress-to-failure distributions, the frequency function is often well represented by the normal distribution function.

Certain statistics obtainable from a normal distribution are used quite extensively in estimating product reliability, yet not a large amount has been written about interval estimates for these statistics.

It is the purpose of this paper to consider the statistic

$$t = \frac{U - \bar{x}}{s}$$

where

U is a constant,

$\bar{x}$ is a sample arithmetic mean and an estimate of the population arithmetic mean,

s is the unbiased estimate of the population standard deviation (σ) obtained from a sample;

and to develop an approximate method for obtaining interval estimates of t. Since $\bar{x}$ and s in the statistic are only estimates and are subject to sampling error, the statistic t will also be subject to sampling errors and clearly will have a sampling distribution. This sampling distribution is referred to as the noncentral t distribution.

Throughout this paper it is assumed that the characteristic in question is normally distributed. Methods are available to check this assumption which range from "by eye" tests to statistical tests with calculable risks of wrong decision. The methods themselves need not be of concern here.

RELIABILITY CALCULATION

Before considering the sampling distribution of t, we will show, by example, how product reliability may be calculated from this statistic.

Consider an electronic component subjected

*Semiconductor Div., Hughes Products, Culver City, Calif.

to extremely heavy constant load conditions until failure. A histogram of the length of life of a large number of these components would appear as shown in Fig. 1, where it is assumed that the distribution shown is well approximated by the normal distribution. Further, it is assumed that the arithmetic mean life is known to be μ and the standard deviation is known to be σ. Thus, there is the normal distribution function shown in Fig. 2.

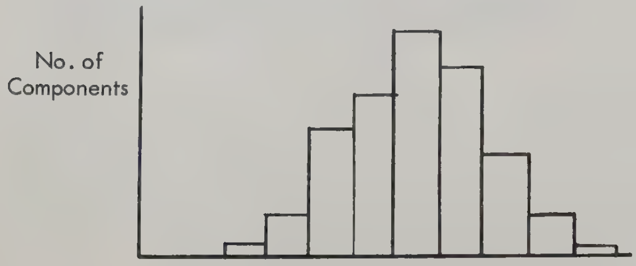


Fig. 1.

Now if a minimum specification (U) has been placed on the length of life of the component, the exact proportion of the components that will fail to meet this specification in the long run can be calculated from

$$z = \frac{U - \mu}{\sigma}, \quad U < \mu ,$$

where z is a standardized normal deviate; from tables of the standardized normal distribution we find

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z} e^{-x^2/2} dx = P_z .$$

P_z then is the proportion of the components that will fail to meet the specification of a length of life of U hours or better. Then $(1 - P_z)$ is the proportion that will meet or better the above requirements. In fact, where the variable x denotes length of life,

$$\Pr (x \geq U) = (1 - P_z) .$$

This says merely that the probability which an individual component will have a length of life greater than the minimum specification U is $1 - P_z$. In keeping with the definition of reliability in common use today,

$$R = 1 - P_z;$$

i.e., the probability of a component performing successfully under the conditions we have set is $1 - P_z$ or R.

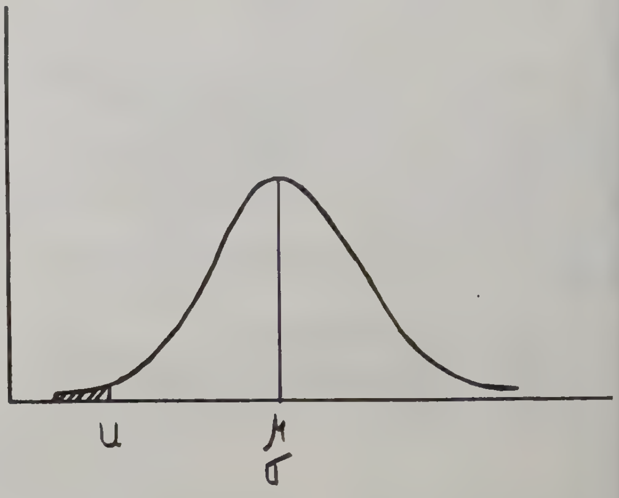


Fig. 2.

Actually, the above example is a trivial case; it was used merely to illustrate the procedure in finding R. The reasons for this are clear. Knowledge of μ and σ (population parameters) implies complete knowledge and enumeration of the population. Thus

1) the exact number of components above U can be counted, and no assumptions about the form of the distribution need be made;
2) rarely, if ever, are μ and σ known.

The more usual case then is the situation in which we have

$\bar{x}$ Estimate of μ
s Estimate of σ

and the statistic becomes

$$t = \frac{U - \bar{x}}{s} .$$

The following section is devoted to finding approximate confidence intervals for the statistic t.

INTERVAL ESTIMATE FOR t

Clearly, the mean value of t (which we will call t') is

$$t' = \frac{U - \mu}{\sigma} .$$

Since t is a nonlinear function of $\bar{x}$ and s, U being a constant, the variance of $t(\sigma_t^2)$ presents

somewhat of a problem. Variances of nonlinear functions are difficult to approximate and sometimes the approximations are not too good, depending, of course, on the function.

In this case, however, where $t = f(\bar{x}, s)$, we can expand the function in a Taylor's series and obtain a good approximation to σ_t^2, provided that we fulfill certain general conditions in addition to the usual conditions necessary for the existence of a Taylor's series. These conditions are as follows.

1) Independence of $\bar{x}$ and s.
2) If we use only the linear terms of the Taylor series, the function should be approximately linear in the region of interest for the independent $(\bar{x}, s)$ variables. The region of interest for the independent variables might be, for example, within ± 3 standard deviations of their mean value. In short, the approximation is best when the deviations from the mean value are small.

In general, it has been shown that for $w = f(x_1, x_2, \ldots x_n)$, the Taylor's series expansion, evaluated at $(\bar{x}_1, \bar{x}_2, \ldots \bar{x}_n)$, yields

$$\sigma_w^2 = \left(\frac{\partial w}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial w}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \ldots + \left(\frac{\partial w}{\partial x_n}\right)^2 \sigma_{x_n}^2 \tag{1}$$

under conditions 1 and 2 listed above.[1] Returning then to our statistic

$$t = \frac{U - \bar{x}}{s}$$

and from (1)

$$\sigma_t^2 = \left(\frac{\partial t}{\partial \bar{x}}\right)^2 \sigma_{\bar{x}}^2 + \left(\frac{\partial t}{\partial s}\right)^2 \sigma_s^2$$

where σ_s^2 is the variance of a standard deviation

$$\frac{\partial t}{\partial \bar{x}} = -\frac{1}{s}, \quad \frac{\partial t}{\partial s} = \frac{-U + \bar{x}}{s^2};$$

but

$$t = \frac{U - \bar{x}}{s}$$

$$-ts = -U + \bar{x}$$

$$\therefore \frac{\partial t}{\partial s} = \frac{-t}{s}.$$

Thus,

$$\sigma_t^2 = \left(\frac{-1}{s}\right)^2 \sigma_{\bar{x}}^2 + \left(\frac{-t}{s}\right)^2 \sigma_s^2;$$

but

$$\sigma_{\bar{x}}^2 \doteq \frac{s^2}{n} \quad \text{and} \quad \sigma_s^2 \doteq \frac{s^2}{2n}$$

$$\sigma_t^2 \doteq \left[\left(\frac{1}{s^2}\right) \cdot \frac{s^2}{n}\right] + \left[\frac{t^2}{s^2} \cdot \frac{s^2}{2n}\right]$$

$$\doteq \left(\frac{1}{n}\right) + \left(\frac{t^2}{2n}\right)$$

$$\doteq \left(\frac{2 + t^2}{2n}\right). \tag{2}$$

Now it can be shown that as n gets very large, the variable t is approximately normally distributed with mean $\left(\frac{U - \mu}{\sigma}\right)$ and variance (2) where t is replaced with t', the true value of the mean. In fact, for most purposes, when $n > 30$, the approximation is satisfactory.

Upon inspection of (2), it may be noticed that the variance of t is dependent upon the magnitude of t. This seems intuitively correct because a proportion below some value U is merely being estimated, and in doing this by using the binomial distribution, it is found that

$$\sigma_P^2 = \left[\frac{P(1 - P)}{n}\right].$$

The variance of a proportion is dependent upon the magnitude of the proportion. Thus, intuitively the variance of t should involve the magnitude of t in some way and it does.

The matter of confidence intervals can now be considered. Of course, the usual type of interval cannot be placed on t; *i.e.*, $t \pm K\sigma_t$ where K is the multiple of the standard deviation that yields a certain confidence coefficient. This would be begging the question, since the σ_t depends in part on the magnitude of our sample estimate t.

Rather, a probability statement of the following nature is made.

$$\Pr\left(t' - K_{\alpha/2}\sqrt{\frac{2 + t'^2}{2n}} \le t \le t' + K_{\alpha/2}\sqrt{\frac{2 + t'^2}{2n}}\right) = 1 - \alpha. \tag{3}$$

Subtracting t' from each member of the inequality and dividing through by

$$\sqrt{\frac{2 + t'^2}{2n}},$$

$$\Pr\left(-K_{\alpha/2} \le \frac{t - t'}{\sqrt{\frac{2 + t'^2}{2n}}} \le K_{\alpha/2}\right) = 1 - \alpha.$$

To find the end points of this interval, we can set

[1] C. A. Bennett and N. L. Franklin, "Statistical Analysis in Chemistry and the Chemical Industry," John Wiley and Sons, Inc., New York, N. Y.; 1954.

$$\frac{t - t'}{\sqrt{\dfrac{2 + t'^2}{2n}}} = \left| K_{\alpha/2} \right| .$$

Omitting the subscript of K for convenience,

$$\frac{(t - t')^2}{\left(\dfrac{2 \alpha t'^2}{2n}\right)} \beta K^2 ,$$

$$2nt^2 - 4 nt t' + 2nt'^2 = 2K^2 + K^2 t'^2 ,$$

$$(2n - K^2) t'^2 - (4nt) t' + 2(nt^2 - K^2) = 0.$$

This is seen to be a quadratic equation in t' with

$$\begin{aligned} a &= 2n - K^2 , \\ b &= -4nt , \\ c &= 2 (nt^2 - K^2) . \end{aligned}$$

Solving, we get

$$t' = \frac{4 nt \pm \sqrt{-8K^4 + 16 nK^2 + 8nK^2 t^2}}{4n - 2t^2} ,$$

Simplification and rearrangement gives

$$t' = \frac{2 nt \pm K\sqrt{-2K^2 + 2n (2 + t^2)}}{2n - K^2} \tag{4}$$

Within the limits of the assumptions which have been made previously, these are the exact confidence limits for t; however, in order to simplify a rather cumbersome formula (4), n is allowed to become large. Under the assumption that n is very large, the denominator becomes essentially equal to 2n.

The quantity under the radical in the numerator becomes approximately

$$2n (2 + t^2) .$$

Then

$$t' \doteq \frac{2 nt}{2 n} \pm \frac{K}{2n} \sqrt{2n (2 + t^2)} ,$$

$$t' \doteq t \pm K\sqrt{\frac{2 + t^2}{2n}} .$$

Replacing the subscript, we have

$$t' \doteq t \pm K_{\alpha/2} \sqrt{\frac{2 + t^2}{2n}} .$$

In short, the usual method of establishing confidence limits is satisfactory, as n gets large.

An excellent discussion of noncentral t distribution applications has been given,[2] as has been a complete exposition of the principles involved in using the distribution in variable sampling plans, and tables of tolerance limits for the normal distribution.[3]

As an illustration of reliability calculation, consider the following practical example.

In a stress to failure test of 100 electrical components, the sample mean length of life and standard deviation were found to be

$$\begin{aligned} \bar{x} &= 4000 \text{ hours}, \\ s &= 360 \text{ hours}, \\ n &= 100 . \end{aligned}$$

Consider a minimum specification set at, say 3300 hours. Then our estimate of t' is

$$t = \frac{3300 - 4000}{360}$$

$$= -1.944 .$$

The normal tables yield a fraction below 3300 of $P = 0.026$. Thus, we can say that the probability of a given component performing successfully is

$$\begin{aligned} R &= 1 - P \\ &= 0.974 . \end{aligned}$$

However, this is probably not the exact value for R and an interval estimate is in order. Using (3) we have, for $\alpha = .05$,

$$\Pr\left(t - K_{\alpha/2}\sqrt{\frac{2 + t^2}{2n}} \le t \le t + K_{\alpha/2}\sqrt{\frac{2+t^2}{2n}}\right)$$

$$= 1 - \alpha$$

since $K_{\alpha/2} = 1.96$,

we have

$$t' = -1.944 \pm 1.96\sqrt{\frac{2 + 3.779}{200}}$$

$$t' = -1.944 \pm .333 .$$

Thus the probability is 0.95 that the interval $t' = -1.611$ to $t' = 2.277$ contains t. Proceeding to the normal tables, we get the fraction of failures lying somewhere between

$$P_{lwr} = 0.0536 \text{ and } P_{upr} = 0.0114$$

$$R_{lwr} = 0.9464 \text{ and } R_{upr} = 0.9886.$$

We are 95 per cent certain that the true reliability lies somewhere between 0.9464 and 0.9886. Mathematically,

$$\Pr (0.946 \le R \le 0.989) = 0.95.$$

[2] N. L. Johnson and B. L. Welch, "Application of the noncentral t distribution," *Biometrika*, vol. 31, pp. 362-389; 1940.

[3] C. Eisenhart, M. W. Hastay, W. A. Wallis, "Techniques of Statistical Analysis," McGraw-Hill Book Co., Inc., New York, N. Y.; 1947.

It should be pointed out that the accuracy of this method in no way justifies carrying as many decimals as has been done. In fact, for very large absolute values of t (high reliability), the approximation is relatively poor, but the absolute error is small. Our estimate of t has smaller variance than the estimate obtained from the binomial distribution.

SELECTION OF SAMPLE SIZE

With the aid of (2), selection of sample size is relatively simple, although not exact. First, an estimate of the reliability of the product must be made. Then, for a certain confidence coefficient $1 - \alpha$, the amount of variability in t which will be tolerate must be tolerated.

For example, an engineer estimates the reliability of a component to be 0.99, but seeks to find the "exact" value.

$$R = 0.99.$$

The normal table gives $t = -2.33$ for $R = 0.99$. Further, the engineer specifies that he wishes to be 95 per cent certain that the estimate t of t' misses t' by no more than ± 10 per cent of t'. In effect, he has now specified the width of the $1.96\,\sigma_t$ limits of the sampling distribution of t.

Thus,

$$1.96\,\sigma_t = .10\,|t'|\,.$$
$$\sigma_t = .05\,|t'|\,.$$

But $|t'|$ is estimated to be 2.33; thus

$$\sigma_t = .05\,(2.33)$$
$$= .116.$$

Substituting in (2)

$$\sigma_t^2 \doteq \frac{2 + t^2}{2n}\ ,$$
$$n \doteq \frac{2 + t^2}{2\sigma_t^2}$$
$$n \doteq \frac{7.33}{.0269}$$
$$n \doteq 273.$$

Of course, it is obvious that if we knew t' we would not have to submit 273 to test, but at least an estimate of t' permits us to predict sample size to a certain extent and, thus, schedule work loads and budget costs.

CONCLUSION

The preceding methods operate to obtain interval estimates for the statistic $t = \frac{U - \bar{x}}{s}$ obtained from a normal distribution and, thus, are applicable to any variable which is normally distributed. The accuracy of the method is subject to fulfillment of the conditions mentioned in the body of this paper. Although only a single tailed problem was cited, a simple extension is two tails, and so are hypotheses concerning t' or the significance of the difference between two t's.

IS ANYTHING NEW IN RELIABILITY?*

W. D. McGuigan†

I accepted the invitation of your chairman to discuss the historical aspects of electronic reliability with mixed emotions. It is quite an honor to open your seminar. On the other hand, I am not quite sure how I came to be regarded as being an authority on the historical aspects of this subject. The last time I had this feeling was about ten years ago. I was trying on a new suit, looking into a three-sectioned mirror, when I saw for the first time that I had a bald spot.

I should like to run over a little history with the purpose of showing that most of the things we are doing for reliability are things we have been doing for a long time, frequently under the same titles. From this, it should be clear that the way we have been fighting reliability we have been solving

*This paper was presented at the First Annual Bay Area Reliability Seminar, Menlo Park, Calif., February 19, 1959.

†Engineering Div., Stanford Research Institute, Menlo Park, Calif.

problems peculiar to the components and systems of the moment, and not really making any lasting contributions to science.

The oldest concern of the reliability hunters has been with the subject of environment. This subject is so old that it actually began about two years before the invention of the audion. Dr. Lee DeForest was experimenting with the conduction and control of electrical current between a Bunsen burner and a pair of ringstands. To shut out variations due to air currents in the room he enclosed the experiment in a tube. Later, seeking complete isolation, he substituted an Edison filament for the burner and sealed the experiment into a light bulb. The birth of electronics in an evacuated glass envelope thus made reliability a congenital problem.

The fight for better components, materials and techniques is an old one too. The work on vacuum tubes at the General Electric Company before and during World War I should stand as a monument to those who would embark on component development programs. One might wonder at the state of their art, however, because while nominally trying to improve vacuum tubes, they invented an extraordinary number of gas filled devices. After nearly 20 years, their leading genius, Dr. Irving Langmuir, even won a Nobel prize for his work on electrical discharges in gases.

Human engineering entered the picture about 1921. Until Harold Elliott made the first gang-tuned Magnavox, it wasn't just anybody who could tune into a signal. This great contribution may have triggered an early, technological counterpart to Parkinson's Law. Instead of simplifying the radioman's task, it simply made it possible for him to handle a greater complexity of equipment.

Perhaps another great step in the reduction of human factors came with the elimination of arc transmitters. These remarkable devices were quenched with pure grain spirits of alcohol which was supplied to most radio shacks in fifty gallon drums. When this was no longer necessary, it is little wonder that engineers redoubled their efforts to seek unusual ways of generating and detecting electromagnetic signals.

An early example of system design for reliability came about 1924. This was the invention of automatic volume control, I believe, by J. V. L. Hogan. Marcus Acheson once characterized this invention as an elephant in the jungle of reliability. Here, for the first time, electronic components could be made to compensate for each other. In one step, the short-term reliability increased perhaps a millionfold.

We should acknowledge another class of contributions which has had much to do with our ability to maintain equipment. The evolution of volt-ohm-meters, oscilloscopes and tube testers, during the 1920's and early 1930's, are examples of essentials which gave us tremendous leverage on the reliability problem.

Fail-safe techniques began to appear in the thirties. While their results were modest compared to the results obtained recently by the AEC, the designers of the low-frequency four-course ranges and, later, the instrument landing systems, were quite aware of their obligations to keep radio beams from drifting off course.

World War II deserves special treatment in a review of reliability. First, the complexity of systems expanded in less than four years by an average factor of perhaps ten. Second, a large number of people without previous exposure to electronics were pressed into its service. Third, and perhaps, more important, the pressure to hurry and simultaneously to change the art led to a tendency to overlook details or at least to compromise them.

During this period, the reliability experts were not electronics engineers. Rather they were a perverted lot, equipped with shake tables, drop tests, salt spray chambers and the like, who, in fact, were unkindly disposed toward electronics. Aside from discouraging engineers, they had no permanent effect on the electronics industry because most of the changes involved things like brackets or platings--things without a permanent role in electronics.

One worthwhile set of postwar concepts dealt with maintenance minimization. Taylor's work on marginal checking in the Whirlwind computer was perhaps the first recognition of the importance of drift rates and design margins. In 1950, Devey, then of ONR, made people aware of the cost of maintenance by setting the now famous maintenance-to-original cost ratio in the range of 10:1. This led to renewed interest in replaceable and throw-away packages. A prominent West Coast Research Institute even had a project to design a fault-finding system for some naval equipment.

After World War II, the very large system appeared. For the first time it was impossible for one talented, well-trained, versatile, omniscient, energetic, personable, persuasive and healthy engineer to understand every aspect of these projects. Computers, planes, and missile systems all got so complicated that the projects outlasted an average of two and a half chief engineers. It then became popular to suspect that the organization rather than the people involved might be causing the trouble.

One organizational example of reliability I like concerns one of the districts of CAA. The engineers in this group had been doing an outstanding

job of maintaining their equipment when suddenly their records showed a fantastic surge in failures. After some research, it was discovered that an enterprising purchasing agent had done them in by purchasing the entire stock of tubes and magnetrons from a war surplus mart.

Reliability papers in the early 1950's frequently fell into stereotypes. The first type was along the theme, "This is a tremendous problem," usually presented by someone of stature in the military. Our commercial friends, meanwhile, started giving speeches along the lines, "Our quality control department reports to 'God'," or "We're confused, but all our data is on IBM."

A very important contribution to reliability, not as well-known to most engineers as I think it should be, is the turn-about occurring in contracting procedures. The Electronic Industries Association committee on Electronic Applications (Reliability) and the DOD's Advisory Group on the Reliability of Electronic Equipment, both under the leadership of Lewis M. Clement, undertook in 1955 to educate some of the conservatives in government on the significance of their contracting procedures.

The general theme was that the military wasn't likely, ever, under existing procedures, to get reliable equipment. The principal arguments in favor of this position were: first, there was no correlation between lowest bids and best reliability; second, separation of responsibility for research, development, production and maintenance placed incentives for getting equipment out of the door rather than for continued operation. The present metamorphosis of this plan is going under the name "value engineering," but I hope none of you will be discouraged by that.

One shouldn't pass the recent history of reliability without mention of the art of error detection and error correction developed by the computer segment of our industry.

To date, the Professional Group on Reliability and Quality Control has published nearly 100 papers. It is worth noting that the most popular single subject (26 papers) concerns the reliability of vacuum tubes. Yet, since January, 1957, there has been only one paper on this subject. As further disillusionment to the tube testers is the recent finding of ARINC that selective testing of modern tubes probably only downgrades the tubes actually used.

The point worth noting is that we have accumulated empirical rather than scientific information about reliability. If we look for things in our past accumulations that will help design engineers, there is relatively little. The factor of ten by which reliability has improved in the past 10 years is far less attributable to our papers on reliability than to the invention of transistors.

If we are looking for lessons out of reliability history, let me suggest the following.

1) Far more research is needed, particularly in instrumentation and statistics.

2) We should ban new systems unless they are likely to improve performance and cut maintenance of an older system by some substantial margin.

3) We should not be carried away by enthusiasts for any large system. Large systems will be down for maintenance, even if they are so-called essential links in our defense system.

4) Above all, don't bother with historical data on electronic reliability. The components, the environments, and the purpose of modern systems are different. Like meteorologists, you should look at yesterday's weather only so long as it helps with tomorrow's prediction.

CONTRACTUAL ASPECTS OF RELIABILITY*

R. W. SMILEY†

The release in April of last year of the so-called Reliability Redbook, "A Proposed Reliability Monitoring Program for the Design, Development and Production of Guided Missile Weapons Systems," has generated considerable interest within Industry about a subject which has long plagued military personnel associated with the procurement of complex weapons. The basic point of argument or discussion, depending upon your point of view, revolves about a single sentence in it, which I quote: "The reliability monitoring program proposed here is based on the premise that reliability is a parameter that can be quantitatively specified, estimated, assessed or measured at predesignated steps or monitoring points of a guided missile weapons system's life cycle and that it can be controlled throughout the phases of design, development, production and major product improvement." If one accepts this premise (its acceptance is almost ideology to statisticians and reliability engineers, but is in the same category as communist ideology to top-level corporation managers), then it follows that it should be possible for the Government to specify by number in a contract a required level of reliability for both development and production. In this way the Government could contractually assure that the high reliability required of these important and complex weapon systems would be attained, and it follows that incentive payments for attainment (or penalty clauses for nonattainment) could be written into those contracts. Then all that would be necessary to implement these contractual requirements is for the contracting agency and the contractor to agree upon a number of monitoring points in the program, establish tests to assess the attained level of reliability of either the design or the hardware, perform the test by either contractor's or service personnel, insert the results of these tests into appropriate formulas, and arrive at a measure of performance upon which the original agreed upon numbers of dollars would then be paid. Such a program I know would be well received by the watch keepers of John Q. Taxpayer's dollars, and it is being given serious consideration by top-level technical and program managers in the procurement agencies of the Department of Defense.

My first thought when I originally received the Redbook was a quotation of Lord Kelvin's which is the watchword of the Bureau of Standards and of our own Bureau of Ordnance measurement standardization program. Eighty years ago, Lord Kelvin said, "When you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the state of science, whatever the matter may be." In all fairness to reliability predictors and assessors, I must concede that in certain instances it is possible to both predict and assess with reasonable statistic validity. Unfortunately, the degree of validity, or more commonly the confidence limits, is in direct ratio to the number of test specimens and in indirect ratio to the number of parts in each specimen. We are considering here the problem of contracting for reliability in highly complex weapons whose productions rates are extremely low and whose unit cost is extremely high. Ballistic missiles today cost two million dollars apiece and up. In spite of a recent newspaper release on the production rate attained by Douglas on the Thor IRBM, I think it is fairly obvious that the aggregate of all missiles being produced today falls far short, for instance, of the 25,000-a-day production rate attained by the automobile industry.

The problem of assessing the attained reliability in a complex weapons system is at least as complex, if not more complex, than the weapons system itself. Most of us are familiar with the definition of reliability published by the Department of Defense Advisory Group on Reliability of Electronic Equipment, which states, "Reliability is the probability that a device, system or equipment will perform satisfactorily for a specified period of time under specified conditions of operation." Translated into the kind of reliability a missile man understands the definition would read, "Reliability means that the missile is ready to go, that it checks out the first time during count-down, that it is launched precisely on time, that it travels

*Presented at First Annual Bay Area Reliability Seminar, Menlo Park, Calif., February 19, 1959.

†Lt. Comdr., U. S. Navy, INSORD, Sunnyvale, Calif.

he prescribed course, reaches the target zone and letonates at the prescribed altitude with the prescribed accuracy.'' Consider just the problem of lefining a successful missile flight. For surface-o-air missiles, for example, there are at least our different definitions of a successful flight, which depend upon the answer to such questions is: ''Is a flight successful or unsuccessful when he missile has deviated from the programmed light *after* the missile has passed the target?'' 'Is a flight successful or unsuccessful when the varhead explodes within the prescribed distance rom the target drone but the drone isn't killed?'' 'Is the flight successful or unsuccessful if the lrone is knocked down even after a highly erratic nidcourse flight?''

There are other problems associated with reiability assessment of missiles besides arriving t a mutually agreeable definition of a successful light.

Missiles are designed to be stored, serviced, andled, checked out and flown by military personel. Although the contractor usually prepares the nstructions governing these many operations, and n many cases provides the ground handling, checkout, storage and launching equipments, it is still xiomatic that the environment in which the missile lives after it leaves the factory is not under he contractor's control. It can be argued that it s the development contractor's responsibility to o design his product that it can successfully withtand normal service environment, but when an xchange of dollars hangs on the performance of a roduct which has been in the customer's hands or a year or more, it is fairly obvious that a conractual debate will ensue on the reason for every ailure.

Another problem stems from the normal conractual arrangements for the supply of the many tems which make up a weapons system. It is rare ven under the ''weapons system manager'' conept that a single prime contractor has total reponsibility for the design and supply of all the quipment in an entire system. Sometimes there s a separate prime for power plants and for heckout equipment; usually there is one for fire ontrol equipment; occasionally even different arts of the missile itself are developed and/or roduced by separate prime contractors. Under hese conditions who is responsible for saying--ndeed, who can say--that a flight would have been uccessful except for manufacturer A's product? ven in those programs where reliability figures re presently released only for information, there s constant bickering among the various contracor and military elements of the procurement, ogistics and operating team as to ''who shot John'' for the relative levels of reliability or for the assignment to the various elements of the causes of unreliability.

Considering these and other difficulties, contracting for reliability along the classic lines, where definition of the requirements and assessment of the delivered item is the basis for payment, is hardly a practical approach to our problem of attaining reliability through contractual action, at least at the level of government prime contracting. When we can afford it, I think everything from bits and pieces to minor assemblies can be purchased to the stated premise of the Reliability Monitoring Program.

Materials at these low levels are characterized by limited numbers of attributes and therefore have limited modes of failures. The subcontracts for these minor assemblies can thus require that the designs or material meet specified reliability goals, and the assessment of the attainment of those goals can normally be made under previously agreed-upon laboratory conditions. Unit costs of items at this level are low enough that we can afford to buy an extra 50 or 100 relays, or valves, or even gyros, to make an assessment within confidence limits that will permit the payment of incentive reward or the assessment of nonachievement penalties. I stress *minor* assemblies as the most complex to be tested because I wonder how many inertial guidance capsules, worth roughly half a million dollars apiece, we can afford to test in laboratories; or how many two hundred thousand dollar ballistic missile solid rocket motors; or, for that matter, how many thirty-five thousand dollar surface-to-air solid rocket boosters. And unless we can afford to test many, the confidence limits are as wide as our CEP.

In brief, then, these are the basic problems connected with specifying in prime missile contracts that a weapons system, or a weapon, must have a certain reliability:

1) The difficulty of arriving at the definition of the attained reliability.

2) The exorbitant cost of making enough tests to assess the degree of attainment.

3) The difficulty of an accurate assignment among elements of the system and organization of attained unreliability.

It might be inferred from the above that the Government can do little to attain the degree of reliability which is required of our present-day complex weapons system. Such is not the case. Furthermore, it can be done through the contractual medium. But the approach is one of engineering instead of one of statistics, a strategic ap-

proach instead of a logistic one. It is the approach that is currently being used in the Polaris program.

It is basic that to attain any goal in a procurement program—economy, high production, schedules, high quality, or a producible design—management attention must be focused on the attainment of the goal. Obtaining and continuing such management attention is the basic premise underlying today's ballistic missile reliability program. Although there is still much to learn about how best to achieve that attention, we are rapidly moving toward an era where the contractor's reliability system and organization will be as well defined and contractually required as his quality control system and organization are today.

We need reliability *today* on the weapons we are developing and building *today*, and we cannot afford to wait until reliability prediction and assessment techniques mature to attain that reliability. Our experience in early missile programs has taught us dramatically and forcefully which elements of a reliability program are most needed to attain reliability.

Briefly and in their approximate order of importance, the principle elements are:

1) A complete design disclosure. It may be surprising that this element is considered to be the most important contribution to reliability. It has, however, been our unfortunate experience that the value attained from the other elements is lost unless the design disclosure is thorough and complete. It does us little good to analyze thoroughly and test a particular relay, to assess and improve its reliability, only to have the vendor change his process or his design without our knowledge. We, therefore, place the highest emphasis on the completeness of drawings, specifications, factory test procedures, field service, test and handling manuals and the other parts of the design disclosure package. Speaking for the Bureau of Ordnance, this package is the means by which the Government attains and retains control of the product, and it is through the medium of the design disclosure package that each level of contractor can control the product he receives from the next lower level.

2) The existence of an element of the contractor's organization, whose primary function is reliability, whose stature is adequate to insure that reliability is heard within the organization, and whose budget and personnel ceiling is large enough to accomplish the required tasks. Just as the Government in the past has delineated a contractor's quality control organization to attain the required level of quality, so we must now delineate a contractor's reliability organization to attain the required level of reliability.

3) A program of reliability testing. The high cost of development for complex weapons makes it imperative that reliability groups integrate into normal development testing as many of the reliability requirements as possible. However, since time is a parameter of more importance to the reliability engineer than to the development engineer, it is also necessary that there be additional testing to destruction, testing in overenvironments, and overtesting in environments which are deliberately chosen to produce failures, in order that we may assess the time-to-failure aspect. It is in this element that we expect to lay the greatest emphasis on designs of experiment, to insure that we get maximum data with the minimum expenditure of precious costly hardware.

4) A separate continuous detailed review of basic designs for reliability. In this day of specialization, one of the anomalies which we frequently face (and one which I am most at a loss to understand) is that in which development personnel resist a separate organization to review designs for reliability. "Reliability is everybody's business" goes the axiom, and its proponents suggest that we can therefore infer that a design coming from a designer whose business is also reliability must be reliable. Unfortunately sad experience has indicated that what is everybody's business is nobody's business, and the job doesn't get done. Just as we expect engineering personnel whose business is production to review designs for producibility, so we must expect engineers familiar with the basic principles of reliability, and having reliability as their prime goal, to review designs for reliability.

5) A feedback system, usually called a Failure and Trouble Report System coupled with an adequate corrective action system which will insure that our experience in the field will result in changes to existing and forthcoming hardware and design.

6) An adequate contractor's and Government quality control organization and system to insure that hardware is produced in accordance with the design requirement. This is not a new element, but it does little good to develop a reliable design and document it so that its requirements are well defined, and then fail to insure that the hardware produced actually meets all those requirements.

At least two of the six elements I have just listed—the requirements for design disclosure, the requirement for quality control organization and system—have been in Bureau of Ordnance contracts for many years. The requirements for the

ther elements are finding their way into newer ontracts and we are slowly learning to define in nore exact terms that for which we wish to conract. The Bureau of Ordnance hopes to have vithin a few months a definitive document specifying these additional elements which we will include as part of the contract requirements. We firmly believe that with proper system and organization, backed up with planning, funds and facilities, it is entirely practical to attain with today's techniques the reliability we need today for today's weapons.

RELIABILITY PREDICTIONS, A CASE HISTORY*

R. A. DAVIS† *AND W. WAHRHAFTIG*‡

The process of predicting reliability is usually reated like the weather; but we found that due to ontractual requirements we had to do something bout it. A description is presented here of the nethods used in predicting the reliability of a iece of complex electronic equipment and an valuation of the results based on field failure ata.

The study to be described was performed main- by the reliability engineers of Philco's Western evelopment Laboratory with an assist from Lockeed's XA Weapons System Reliability Engineerng Department. The data used for prediction ere supplied through the courtesy of RCA at ape Canaveral, Fla. The data used in evaluation ere taken from Trouble and Failure Reports sed in the field by operating personnel.

One of the major pieces of ground equipment of e early part of the XA Weapons System is the Verlort'' Radar. This is a modification of the lod II radar used at AFMTC. It in turn is a modication of the 584 radar. In order to evaluate the eliability of the Weapons System, it is necessary study its components. This paper shows how e very long-range tracking radar, the Verlort, as studied.

For prediction, the approach taken was first to valuate the existing Mod II radar and extrapolate e results to the new system. To make this evalation, the operating logs of the Mod II radars ere examined. It was found that about half of ese had sufficient data to make statements about ean time to failure and mean time to repair.

*Presented at First Annual Bay Area Reliability minar, Menlo Park, Calif., February 19, 1959.
†Western Dev. Lab., Philco Corp., Palo Alto, Calif.
‡Missiles and Space Div., Lockheed Aircraft Corp., nnyvale, Calif.

It should be stated at this time that in order to predict reliability using any model, the important statistics to be determined are mean time to failure and mean time to repair.

The data gave:

Total operating time = 8388 hours
Number of failures = 343.

This yields a mean time between failures, $\overline{T}$, of 24.4 hours. Mean time to repair, t, was computed by averaging the individual times to reapir, and was found to be 1.85 hours.

Fig. 1 shows time to failure plotted with a 90 per cent confidence interval for each point. This

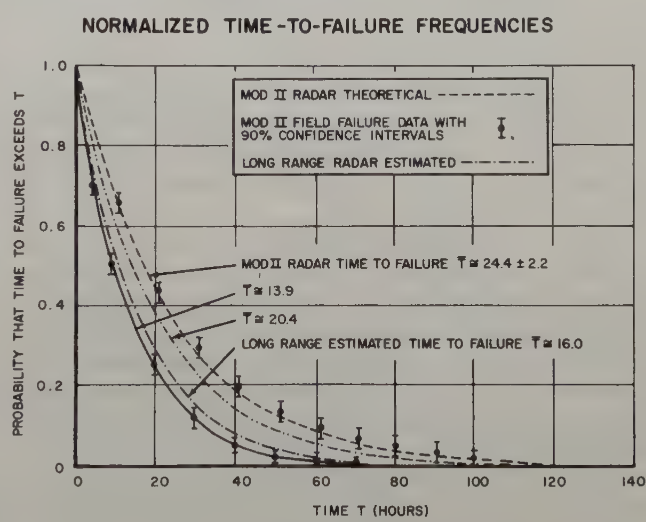


Fig. 1.

also demonstrates how well the data fit the frequently assumed exponential model for time to failure. In Fig. 2, time to repair is similarly

plotted. It should be noted that this, too, follows the exponential model.

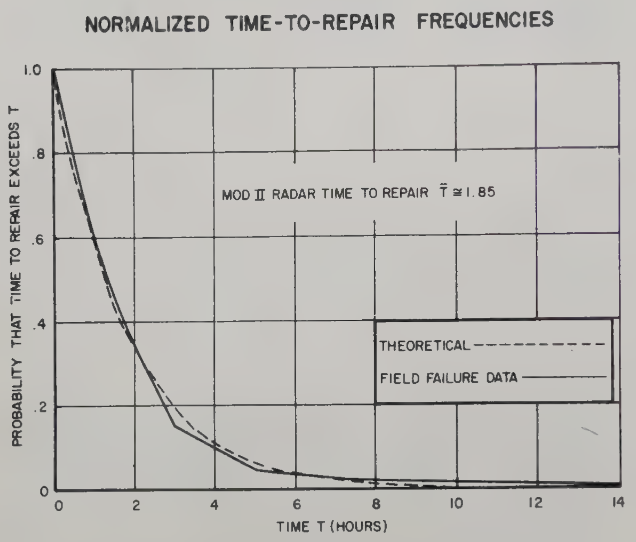


Fig. 2.

To evaluate the reliability of the new radar it was necessary to determine how it differed from the existing model. The Mod II has 60 major components. The Verlort has 88. These 88 fall into three categories: existing, the same as used in the Mod II; modified, similar to those used but with varying degrees of change; and new, not used in the Mod II.

The most direct method for extrapolating the field failure data of the AFMTC radars to the very long-range system is to assume that the 88 units of the Verlort system have the same average failure rate per component as that found on the 60 units of the Mod II.

Such an assumption appears to be valid. In examining circuits of both the old and the new components, the application of parts, the circuit configuration, and packaging employed were found to be similar (some circuits were actually analyzed using the stress analysis techniques described by RCA). The functions to be performed by the new circuits were not radically different from the original circuitry, and so no attempt to push the "state of the art" had to be reconciled. The circuits that were modified were for the most part changed to improve reliability.

Mod II failure rate

$$= \frac{1}{\bar{T}} = \frac{1}{24.4} = 0.04098/\text{hour}$$

Average component failure rate

$$= \frac{0.04098}{60} = 0.000683/\text{hour}$$

Long-range radar failure rate

$$= 0.000683 \times 88 = 0.06010/\text{hour}$$

Long-range radar mean time to failure

$$= \frac{1}{0.0610} = 16.6 \text{ hours}$$

As a check, another method was utilized in making the extrapolation. A component layout drawing was obtained from Reeves Instrument Company. It showed the block layout of the very long-range radar and general information regarding the parts within each component in the block diagram. Failure rates were then assigned to the Verlort by dividing the 88 components shown on the Reeves drawing into the three categories—existing, modified, and new. The estimated mean times between failure were computed by comparing the existing component complexities in terms of their failure rates with the knowledge available about components being modified and new components. Summation of failure rates for the long-range radar yields the system mean time between failures shown below.

The failure rates are a summation of the reciprocals of the estimated mean time between failures.

	Number	*Failure Rate*
Unchanged components	46	0.01819651 per hour
25 used once		
5 used more than once		
Modified components	20	0.02168838
New components	22	0.02318780
TOTALS	88	0.06307269 per hour

Mean time between failure

$$= \frac{1}{0.06307269} = 15.85 \text{ hours.}$$

The two methods used above yielding mean time between failures of 15.85 and 16.6 hours indicate that 16.0 hours is a good approximation to be used.

Both procedures required that for both systems the population is composed of items of similar construction with similar modes of failure and

that over a long period of time the MTBF of such complex systems approaches an asymptotic value. The model assumed for the time to failure is not important. Although our data as indicated on Fig. 1 fitted the exponential model, the same assumption of homogeneity would apply to any other model. Fig. 3 shows the percentage part replacements for both the Mod II and the long-range radar, and supports the assumptions.

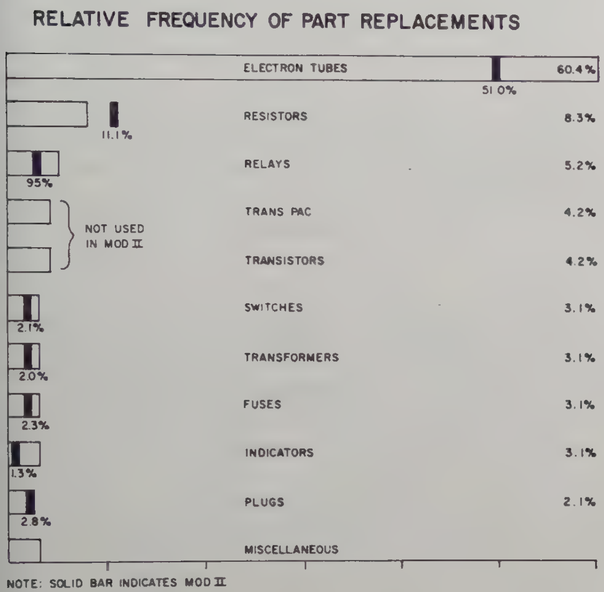


Fig. 3.

As for mean time to repair, this is mainly a function of troubleshooting procedures and personnel training. Since we anticipated these to be about the same as those used at AFMTC, we assumed the same mean time to repair—1.85 hours.

Let us now see how good these estimates were. One hundred and thirty-four Trouble and Failure Reports from operating Verlort radars were analyzed. These reported 56 "operating failures" in 1140 hours of operating time. This yields a mean time to failure of 20.4 hours. The 90 per cent confidence limits on this place the estimate between 16.6 hours and 25.8 hours. This means time to failure is represented by the second curve on Fig. 1.[1] The mean time to repair was found to be 1.3 hours. It can be seen that the estimates are quite close to the values found in actual operation. (If all the failures reported from the field had been included, $\widetilde{T} \cong 13.9$ with limits of 11.7 and 16.9 Shown on the lower curve of Fig. 1.)

Two points should be mentioned briefly. First, a great deal of benefit is obtained from gathering data of this sort besides its use in determining reliability statistics. From both the Mod II logs and the Trouble and Failure Reports valuable information was obtained for pinpointing trouble areas, for establishing a preventive maintenance program, and for help in determining spares.

Second, it should be noted that nowhere above was reliability calculated. Reliability, or for equipment of this type availability, is a complex function of the operating requirements. As a simple example to illustrate a difference between the two radars, the following model might be used. It makes these assumptions:

1) The system is either in use or being repaired.
2) If the system is in a failed state at the outset of its use period or fails during the period, it will not be repaired before the end of the period.
3) The time at which the system is required is independent of whether it will operate.
4) The operating time, T, is 16 minutes.

In this case, reliability, R, is given by:

$$R = \frac{\overline{T}}{\overline{T} + \overline{t}} e^{-T/\overline{T}}$$

For the Mod II R = 0.91, for the Verlort, R = 0.88.

The foregoing illustrates that for large systems we are able to predict, with a fair degree of accuracy, those statistics required to evaluate the equipment from a reliability viewpoint. In this case, we were fortunate to have available a great deal of data from which to start. However, in most cases, sufficient data exist on similar equipment and/or similar operating requirements for adequate predictions to be made.

[1] The confidence interval is based on an assumption of an exponential distribution that Mod II experience indicates.

REFERENCES

[1] "Verlort Radar: Field Failure Analysis," attached to Philco memo from Quality Assurance Group to Distribution on Analyses of "Trouble and Failure Reports" for Verlort Radar, Western Dev. Labs., Philco Corp., Palo Alto, Calif., January 22, 1959.
[2] "Reliability of Long Range Tracking Radar," Missiles Systems Div., Lockheed Aircraft Corp. Rept. No. LMSD-6152; September 5, 1958.
[3] D. Bentley, "Determination of Availability," Western Dev. Labs., Philco Corp., Palo Alto, Calif., Tech. Rept. No. 1084; August, 1958.
[4] "Reliability Stress Analysis for Electronic Equipment," RCA, Cape Canaveral, Fla., Tech. Rept. No. 1100; November 28, 1956.
[5] Shell and Saar, "Mod II Radar—Preventive Mainte-

nance Program," 2nd Edition, April, 1957, Installation and Maintenance Range Operations, RCA Missile Test Project—PAFB, Florida.
[6] "AFMTC Tracking Radar Mod II—Vol. II Theory of Operation, Trouble Shooting and Repair," Reeves Instrument Corp., (Confidential).
[7] "Handbook of Preferred Circuits," Navy Aeronautical Electronic Equipment—(Nav Aer 16-1-519). Prepared by National Bureau of Commerce for Bureau of Aeronautics, Dept. of Navy, September, 1955.
[8] Pearson and Hartley, "Biometrika Tables for Statisticians," vol. 1, Table 40; 1956.
[9] H. Cramer, "Mathematical Methods of Statistics," Princeton Univ. Press, Princeton, N. J.; 1945.
[10] N. S. Hawley and J. H. Rowland, "A Mathematical Model for Reliability," Missiles System Div., Lockheed Aircraft Corp., Sunnyvale, Calif.; Rept. No. LMSD-2278; September, 1957.

CONTRACTOR MANAGEMENT LOOKS AT RELIABILITY PROGRAM ACTIVITIES*

W. B. LABERGE†

Most of us who are here today live several quite different lives. For a certain portion of each day and of each week, we engage in a business conducted in a business environment with its specialized set of objectives. Quite separate from this, in the remainder of our lives, we are engaged in social activities and a pursuance of objectives set by our social standards, perhaps towards objectives quite different from those of our business life. Through periods of our adolescence, educational development, and throughout our adult life, we have modified the way by which we live in this social environment, adapted to it, and by adapting to it are permitted achievement of our objectives. Within this common social climate, we are quite able to have individual sets of objectives and individual moral and ethical codes. Despite these individual differences, however, we must admit to being in a common society. This common society having a reasonably diverse set of component parts is still regulated to achieve the common good.

Although we do not, in detail, always know the precise direction in which we should go, or the detailed actions which we should take, to achieve these goals we do not here need a lecture about the social environment in which we live. We have lived in that environment and we have become able to know and understand it.

*This paper was presented at the First Annual Bay Area Reliability Seminar, Menlo Park, Calif. February 19, 1959.

†Western Dev. Lab., Philco Corp., Palo Alto, Calif.

What I would like to do with you for a little while is to present a discussion of the business environment in which you live, which perhaps is not nearly as well known to you as your social environment, and discuss with you a few of the constraints and restrictions and opportunities which form a part of this environment. As engineers engaged in reliability activities, you form the part of a quite new area of engineering activity.

This activity is deposited in an existing business society whose standards have been unchanged for many years. It is as if you were the equivalent of a minority group placed in a staid suburban area which of its own would not choose to have such an addition. If you accept this equivalence to a minority group, it must then be recognized that it is your responsibility to overtly strive to fit into the business environment. For without your efforts to integrate into the work towards the common good, this business society will do what its social counterpart would do; isolate and wall you from it.

The reason for this reaction to a new group is the same kind of fear which underlies the treatment of social minorities, namely a lack of understanding of a minority's function and a worry that it will take over or retard something which is not theirs.

Therefore, let me make a first point: a positive effort is required by a reliability group to assure all members of the business society that reliability groups can and will effectively work toward the common business goal in a harmonious integrated way. This is quite crucial. It is a requirement of each man in a reliability program as well as the

equirement of its most senior people. If this ef-ort is not made, you will be isolated and walled-p by some mechanism or other. By a series f organizations and reorganizations, you can be-ome further and further detached physically and rganizationally from what is really the interest f the business of which you are a part.

If one can achieve social acceptance now, then ne can turn to the next most important question, hat of organization and responsibility assignment. One realizes that there has always been a desire nd, in fact, a requirement for high quality in quipment produced by the manufacturer. There as always been some mechanism established ithin the business environment to monitor, upervise, and encourage reliability of operation f equipment. What is new and different about eliability activities has been caused by the tre-nendous growth in the technological requirements f equipments. What had before been relatively traight-forward mechanical arrangements of arts now have become very complex assemblies f very complex individual parts. Furthermore, fter grouping of these parts into assemblies, hese assemblies have been further grouped into ery major collections of assemblies called "systems." The net result is that the individual ontracted item is much more immense than it sed to be in its total assemblage of parts. No onger does a straighforward workman-like job ssembling parts permit assurance that the major ystem resulting from these parts is satisfactory. This has led to a business emphasis on system esign, and with it an emphasis on system relia-ility, and through it an emphasis on individual omponent part reliability.

If one admits reliability programs have hanged, we can look at how management reacts o this change. First, let me say that it is obvious o you and obvious to any competent management ctivity that the engineering force, both in its esign and in its reliability facets, is the back-one of an R and D organization. Without a re-iable product, there is no reason for existence f a research and development or an engineering ctivity. However, these are not the only parts, or the only important parts of a business organ-zation. Although these other parts do not neces-arily require engineering education or skills in heir proper execution, they do require a high evel of competence and have a very important mpact on the business itself. The combined arts of this business strive to execute their noral responsibility. That moral responsibility hich any properly conducted R and D activity as is to show a reasonable financial profit and to place that organization, by expansion of its facilities and capabilities, in such a position as to ensure its continued growth and ability to provide products and services purchasable by its customers. It needs to be clearly recognized and clearly understood that not only is it the way things happen in a business society, but that it is the moral and ethical responsibility of the management of any organization to show a profit for those who have invested capital in that enterprise. Each of you who invest in other enterprises expect to see these enterprises grow and prosper and a reasonable return to be made upon the investment which you have made. So also the business in which you are now working is required by the same moral and ethical responsibility to provide a profit to its supporters.

What this means, therefore, is that the maximum economy of operation must be exercised in order that one can most straightforwardly and most economically pursue the objectives of this business enterprise. Surely the engineering staff does not wish to consider itself as accountants or controllers or plant facility people but similarly the plant facility personnel, accountants and the controllers frequently do not wish to be considered engineers. Each are separate and integral portions of the business which is being conducted. The function of a business management is to evaluate the individual contribution of each of the integral parts of the business operation and to provide a management structure which places them in a proper line position to ensure correct emphasis on these individual parts.

With respect to reliability, proper organization is perhaps one of the most challenging problems which a management can have. Within each of the major structural elements of a business organization, the influence of reliability activities is felt. There is no clean-cut separation of reliability in the engineering, plant operations or the fiscal or the production areas. An example of this is perhaps the obvious one of a quality control program. It must effect its influence not only in the engineering development of a product, but on the methods for high-scale production. It must exert its influence upon the purchasing side of the house by proper selection of component parts vendors, and also it inevitably must affect the controller's office through the cost that the business must bear to support quality control.

Before one speaks much further about the management problems associated with reliability, it is necessary to define what management wishes from a reliability program. Perhaps within the context of this paper, these objectives can simply

be stated into two parts. First reliability programs must participate in the line responsibility of an engineering department charged with the development and fabrication of a product for a customer. Secondly, reliability programs must provide an independent audit to management, reporting the prognosis of success of a given program during the early course of its engineering development.

If one admits that the two requirements exist for reliability activities, that of line and audit, then reliability must be organized both in a line capacity and in a staff capacity. These two obviously incompatible simultaneous requirements present a problem to management. They also present a problem to you as a member of a reliability program, for your personal success and that of your group are measured by your contribution. This contribution is frequently significantly modified by the organizational structure within which reliability works.

Therefore, a second point I bring you is that as members of a reliability program, you must ensure by discussion with the business management that your organizational position within the company structure permits achievement of the aims of the reliability program.

The organization of reliability programs in one company can differ from the organization in another, since company requirements differ in many ways among each other. A claim cannot be made that one way of organization is manifestly better than another. Nor does one claim that an organization must be permanent. In fact, it must change as the conditions defining it change. However, as in your social world, so also in your business world, before you can act you must be in a position to act.

If the preceding recommendations were outward in their relations to others, the next few will tend to be more inward looking. Reliability programs must recognize that management views their activities in a way not really very different from the way it views the activities of the accounting, production and purchasing departments. The overriding questions always asked are: Is the service necessary? If so, what size service is required? Is the service provided competently performed? Is it worth the cost it sustains? These are essential management questions and are inevitably asked.

Therefore, one may make the next points directly. Reliability programs must directly address themselves to the problems of the business concern for which they work. Perhaps this is obvious to you, but it is unfortunate how frequently this point is ignored. To reiterate, the normal reliability responsibility is to work on the problem assigned it.

Many too many times, Parkinson's law operating on otherwise intelligent engineers creates a wholly-closed intellectual world where masses of paper work are distributed to those within that world but out of which no useful output appears. A second frequently seen occurrence is that of the reliability group which has become so erudite that no one can understand it. As it becomes so knowledgeable on the details of its subject, it ceases to care whether any one of dissimilar intellectual background can understand it. Finally, it feeds only on the praise it internally generates for itself. Statistics for statistics sake has a place, as does any basic research program, but if you accept your check on the basis of work on a specific problem, then output is the major measure of your worth in this society you joined.

Immediately after groping to see that there is an output, a management must try and ascertain if it is a competent output. It is surely so that quite some number of management leaders are themselves not competent to judge. However, they must so judge. This is not unique to reliability; each major function of a business is likewise judged on the basis of somewhat inexact standards. One of these inexact standards is one's view of the competency of the people involved. To be competent as a group, one must employ competent people. A reliability department or an engineering department may obtain excellent people or it may acquire poor ones. Good people tend to attract good people and poor people to attract poor people. So progressively, the competence of a group increases or decreases but seldom stays static. Inevitably, on a day by day basis, management views the quality of a program by the quality of the people it sees in that program. Therefore, a next major point to be made is that "a Reliability group must be competent and appear competent to survive."

So far, in order, management recommendations to reliability activities might be to:

1) make themselves a part of the business they are in;
2) see to it that they are organized so as to permit effectiveness;
3) work on the assigned problem to them; and
4) be competent and demonstrate competence.

Lastly, the most important point, reliability must provide a useful output. In this business environment of which one has spoken, it is the output which is sold. If the return by sale ex-

ceeds the cost of the product or service, a profit is made. If not, a loss accrues. The long-term objective of a business must be to show a suitable profit. If a reliability program, by useful output, does not contribute a measure equivalent to its cost, then major remedial action is incumbent. Therefore, this strong emphasis on this fifth and last point which is, to reiterate, a reliability program must above all have an acknowledged useful output.

By now, if you are still rational and not emotionally overwrought by the directness of this so-called management view, you will have noticed perhaps one obvious point. What has been said about reliability programs can be said in almost an identical manner about each and every other of the major departments of a business. This is as it should be. You are a part of this business society, and if accepted by it, should be expected to be judged by that society's standards.

In conclusion, perhaps one final statement can be made. As the speaker gives management's view, simultaneously you are given a view of management. In this business society of which one speaks, you must live with management as it must live with you. Each has a common aim, the interesting challenge is to see how each can help the other, so that cooperatively they can achieve their common aim.

DESIGN INFORMATION INTERCHANGE AMONG CO-CONTRACTORS*

MARTIN BARBE†

INTRODUCTION

My paper will describe some techniques to facilitate intercontractor reliability data exchange, and will suggest procedures to increase the utilization of such data after receipt. It will also outline the normal activities of the Space Technology Laboratories and the Ballistic Missile Division of the Air Force, in these general areas.

I would like to establish the boundary limits of my discussion by identifying it as dealing with the flow of *information between* contractors, rather than the flow of instruction from a prime contractor or technical direction agency, as the title in the program might imply. There are areas where the two functions overlap, especially where informative material is available to the prime contractor, and desired by the associate contractors; and neither wishes it transmitted as contractual direction. However, my topic does not primarily deal with this informing or educating of subcontractors, although some of the procedures described can be used to facilitate this action.

In late 1957, STL/BMD instituted a system for the interchange of nonproprietary data on the testing of component parts among the Titan contractors. Fig. 1 may clarify the agencies referred to.

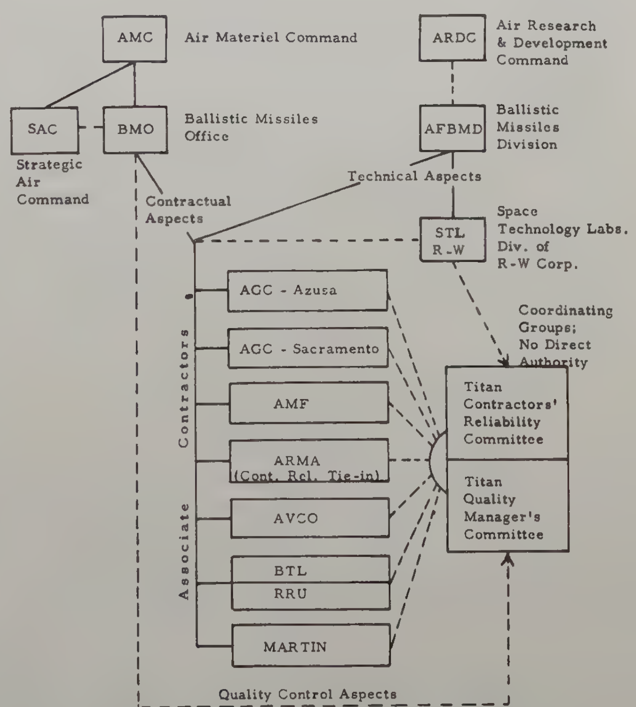


Fig. 1—General functions of participants in Titan Missile Program, as they tie in with the reliability activities described.

*This paper was presented at the First Annual Bay Area Reliability Seminar, Menlo Park, Calif., February 19, 1959.

†Space Technology Labs., Rano-Wooldridge Corp., Los Angeles 45, Calif.

In mid-1958 this was extended to cover the airframe contractors for Titan, Atlas and Thor; and interchange set up with NOL (Corona) and ABMA (Redstone), plus recent initiation of a specialized interchange between Air Force propulsion contractors.

The interchange between Titan contractors included not only test reports, but also general reliability procedural plans, reports on design techniques and manufacturing processes, procurement specifications, and educational bulletins and films on reliability. (The interchanges between different programs and between Services have been more restricted to component part test reports.) It is obvious that such *general* documents are of maximum usefulness near the inception of a program and their utility decreases as procedures are established. We have therefore concentrated on reports of laboratory tests, run under controlled and recorded conditions of load and environment, on equipment which is completely identified and obtainable by other users.

It is expected that the cross-program exchanges by contractors on similar portions of the weapon system will prove far more productive than the "single program" exchanges, even though smaller parts and GSE components may be common to several contractors on one program.

Although the reports may not describe tests conducted under exactly the environments or to the limits on which positive information is desired, they can often point out modes of failure or performance to investigate, which allows a much shorter verification test. Such data can be of help to a contractor who has only a minor portion of his equipment of a type on which another contractor has been able to afford considerable specialization.

Individual reports of field failures are not transmitted between contractors, although information on serious problem areas may be.

MECHANICS AND MONITORING OF INTERCHANGE FLOW

The interchange is initiated by a contractor-to-contractor exchange of list of new reports, available on direct request. (See Fig. 2.) STL monitors this activity by information copy only, and give such assistance as organizational meetings, issuance of consolidated listings, standard forms, etc.

In establishing this activity, STL has followed these steps:

1) Obtain each contractor's top management approval to the identifying of a single individual as focal point or "coordinator" for interchanges, to correspond directly with similar coordinators at other contractors.
2) Promote cooperative action between coordinators by personal contact and direction where required.
3) Encourage monthly exchange of *lists* of experimental reports available on request. (Requests would follow after examination of lists for titles of interest.)
4) Encourage utilization of listings *within* the contractor's plants through periodic STL/BMD issuance of *subject oriented* consolidated listings of all current reports available to all contractors.
5) Monitor for prompt cooperation, through copies of all requests and transmittals.

CONTRACTUAL ASPECTS

The initial program was established by STL on a purely voluntary basis, and this noncontractual aspect was later emphasized by a BMD letter to contractor top management in the Spring of 1958. Almost 100 reports per year have been volunteered by each participating contractor so that, while contractual direction of the production and offering of these reports might be desirable in the future, it has not been imposed as yet.

There are and have been many efforts made in this area of data exchange. I will describe some of the current or pertinent activities with which I am familiar.

Electronic Industries Association Activity

At present, any action on Qualification Data Interchange by the QA-1 Group is at a standstill. Likewise, the M-5 group of the EIA is no longer contemplating data exchange. (However, the system developed for environmental-level coding on punch cards may be of interest and information on this and on the choice of environments and intensities is available on request.)

Aircraft Industries Association Guided Missile Committee

The latest information received is that suggestions in the area of the data interchange are receiving specific attention by the legal staffs of AIA members. At present the Aircraft Research & Test Committee made up of the chiefs of the test labs, has a limited informal exchange of test reports, on a personal basis.

Date: 9-6-57
File No.: 47. -16

TITAN PROGRAM

RELIABILITY INFORMATION INTERCHANGE

☒ Transmittal
☐ Request

FROM	TO	INFO. COPIES		
			Aerojet — Azusu	B. Wilner
			Aerojet — Sac	J. J. Peterson
			Arma	E. Dertinger
			Avco	J. Leary
			BTL	T. Winternitz
			MARTIN	A. Rhoads
			RRU	G. Raymond
			R-W	Proj. Eng.

Subject: Listing of Recent Test Data

References: none

Enclosures: none

The following Environmental Laboratory Test Reports are available to all interested parties:

No.	
33	Barium Titanate Accelerometers - Gulton
37	Accelerometer Transmitters - Giannini
28	5 Volt Mercury Cell Battery - Mallory
55	30 Volt Primary Battery - Yardney
51	Self-amplifying Accelerometers - Gulton
201	Test Report of the Rapid Throwover Relay MS 25024-1 and MS-25035-1 Relay
19	Accelerometers - Statham
26	Slot-Antenna - Radiation Inc.
75	Battery Cells - Nicad
29	Battery - Saft Voltabloc
71	50 Hour Salt Spray of Beryllium Specimen
233	Life tests on 2N135, 2N136, 2N137, 2N123, 4JD1A17, 2N45, SB100
255	Effect of 2 kinds of flux on printed circuit boards
279	Effect of temperature, humidity, temperature shock and salt spray on Amphenol No. 111394 connectors
E-1-B	Silver-Zinc Battery, Comparison of AMF and Cook Company Battery
E-18.1	Rotary Frequency Converter, Functional Tests

UNLESS OTHERWISE NEGOTIATED IN ADVANCE, ANY ACTIONS GENERATED BY THIS TRANSMITTAL ARE CONSIDERED VOLUNTARY IMPLEMENTATION OF EXISTING CONTRACTURAL OBLIGATIONS TO SUPPLY THE U. S. GOVERNMENT WITH RELIABLE WEAPONS SYSTEM ELEMENTS, AND DO NOT OBLIGATE THE REQUESTOR FOR ANY EXPENDITURES INCURRED.

ANY INFORMATION TRANSMITTED IS INTENDED TO PROMOTE IDEA INTERCHANGE AND EVALUATION AMONG GROUPS INVOLVED IN THE GUIDED MISSILE PROGRAM IN THE NATIONAL INTREST, AND ITS TRANSMITTAL DOES NOT IMPLY VERIFICATION OR ENDORSEMENT. BY SUCH ACTION, THE TRANSMITTER ASSUMES NO LIABILITY TO ANY PATENT OWNER NOR ANY RESPONSIBILITY OR OBLIGATION WHATSOEVER TO PARTIES ADOPTING ANY PRODUCTS OR PRACTICES NOR ANY RESPONSIBILITY OR LIABILITY FOR COMMENTS ABOUT ANY PRODUCTS OR PROCESSES. APPROVED PRODUCTS AND PRACTICES ARE THOSE DEEMED SATISFACTORY FOR THE PARTICULAR PURPOSES AND STANDARDS OF THE GUIDED MISSILE PROGRAM; AND NO ATTEMPT HAS BEEN MADE TO TEST OR EVALUATE ALL PRODUCTS AND PRACTICES WHICH MIGHT PROVE SATISFACTORY.

Fig. 2—List of lab test reports for exchange between contractors in Titan Missile Program.

Project 2, Task B of the Electronic Reliability Panel of AIA

This project is now drafting plans to bring before the parent committee, to follow on from the previous efforts of the AIA *Reliability* Sub-Committee. Present thinking of the Project is to encourage a series of "networks" for interchange among special interest groups, such as the ballistic missile contractors; and later possibly to merge these into one.

"HELPER" Program of Inland Testing Labs

The proposed program, involving large-scale tests under a wide variety of environments, conducted by a central test Lab but financed by a number of contributing corporations, has now been abandoned.

Battelle Memorial Institute Program

Battelle proposed a cooperative data analysis and summarization service to a number of large concerns, in the fall of 1958, and now states that they have the required ten subscribers, at $20,000 each, This is an extensive activity involving review by Battelle engineers of contractors' raw data at their plants, and the compiling of this into periodic state-of-the-art reports on particular components. If successful as planned, it might eventually supersede the direct interchange of full report tests; but, for the present, its stage of development and longer time-cycle seems to make it aim at a somewhat different requirement.

Activity by the Office of the Assistant Secretary of Defense

In the Fall, 1958, Mr. E. J. Nucci instituted an Ad Hock Study Group on Electronic Parts Specification, Management and Reliability, which included the topic of centralizing information on component performance. However, until the present, their work has been largely preliminary, although they are very interested in assisting any activities mounted in this line by non-Government groups.

The exchange system described works suitably enough for up to about seven contractors, but, if expanded, the problems of communication *from* all contractors to all contractors become cumbersome.

The Air Force is cooperating with ABMA and INSORD (Sunnyvale) in attempting to arrange an inter-Service exchange of data between all ballistic missile contractors. One plan being discussed involves a central disseminating agency, working with ASTIA, and utilizing microcards. If adopted and proven satisfactory, this should increase the availability and utility of the data to the end user. The procedure might also assist by relieving the technical people participating in an interchange system of the burden of handling the routine transmittals, leaving their time for follow-on inquiries or topics not suitable for a standardized transmittal.

Specific contractual coverage of interchange (as now called out on one Ballistic Missile contract) might facilitate the activity, and this point is being considered on new contracts.

However, various development in data exchange such as punched-card coding of the experiment description, centralization of this activity on any national basis, standardization of format, summary, procedures or even requirements of component test reporting; all these are quite conditional on some positive demonstration that the information transmitted will actually be believed and demonstrably utilized to benefit the national defense effort.

HUMAN FACTORS IN THE ATTAINMENT OF RELIABILITY*

R. S. LINCOLN†

Human factors influence the reliability of complex equipment in many ways. Such a statement should surprise no one since human beings are highly involved in the design, fabrication, operation and maintenance of every machine that we have. Despite the general agreement on this statement, systematic attention to the human factors problem has only recently been extensively applied. In this discussion I would like to concentrate on two different (but related) aspects of the human factors problem and briefly describe some useful remedies. The distinction between the two aspects results from a concern for the operators of the equipment as contrasted to a concern for the design and reliability engineers who interact in designing the equipment.

In dealing with the operation of equipment we will call upon an activity currently known as "human engineering" or "engineering phychology." In our concern for equipment designers we will depend, at least in part, upon a body of theory and research known as "group dynamics." For convenience I will identify problems involving operators as man-machine problems and problems involving designers as man-man problems. The discussion of man-machine problems will be concerned with recommendations regarding the reduction of errors in the operation of equipment. The discussion of man-man problems will be concerned with improving the acceptance of those recommendations by design engineers.

MAN-MACHINE PROBLEMS

Serious interest in man-machine problems is usually identified with World War II during which the design of complex equipment began to place unusual demands on the human operators of that equipment. Plagued by their human limitations, operators were often incapable of consistently performing their tasks in a safe manner. Numerous examples demonstrate that aircraft designers cannot take anything for granted when a human being is involved. During the last war, for example, one type of airplane was made with a large door that permitted access to the push-pull rods controlling the elevators. One pilot, not realizing what the rods were for, strapped a suitcase to them before taking off on a flight. After take-off he was lucky enough to be able to circle and land without crashing. Upon landing he complained bitterly about the difficulty he experienced in moving his elevators, until his own mistake was discovered.

Even in the Civil War, with the relatively simple equipment then in use, there were frequent errors that could be related to equipment design. It has been reported that after the Battle of Gettysburg, several thousand abandoned muskets were picked up. About one-half of these muskets had two charges in them, and more than one-quarter of the muskets had from three to ten charges rammed into their barrels. Apparently, under stress the relatively simple procedure of loading and firing a musket became overpowering, just when efficient operation was most critical.

Stories such as these have led some persons to say that equipment must be designed so as to be "idiot proof." I object to this term, however, because in using it an important point is overlooked. When equipment is not designed to prevent them, or at least reduce their likelihood, errors will still be made by the most highly trained, intelligent of personnel operating with the best of intentions. How many people, for example, have never turned on the wrong burner of a stove or stalled their automobile engine because they forgot to shift gears after experiencing some stressful driving condition?

Errors in the operation of equipment interfere with the attainment of reliability in the same way as component failures. Their ultimate effect may well be a failure of the mission itself. With a man-machine system, therefore, human errors are an appropriate subject for the recommendations of reliability personnel.

In order to determine where the human engineer should direct his attention, we should first identify the types of error to which human operators are prone.

TYPES OF ERRORS

For convenience, the kinds of errors made in operating equipment may be classified as shown in Table I.

*This paper was presented at the First Annual Bay Area Reliability Seminar, Menlo Park, Calif., February 19, 1959.

†Reliability Engrg. Dept., Missiles and Space Div., Lockheed Aircraft Corp., Sunnyvale, Calif.

TABLE I

HUMAN ERRORS RELATED TO EQUIPMENT DESIGN

1) Errors of omission
   a) Errors of memory
   b) Erros of attention
2) Errors of commission
   a) Errors of identification
   b) Errors of interpretation
   c) Errors of operation

Errors of omission are errors in which the operator plays a passive role. They result from things the operator does not do. Errors of commission, in contrast, are errors in which the operator is more active. They result from things the operator does do. The errors in each category will be described in the following discussion together with specific examples and possible recommendations.

Errors of Memory

Errors of memory occur when the operator forgets to carry out a task or forgets the sequence in which an operation is supposed to be performed. The latter error is especially likely when the operator must work with equipment panels containing numerous displays and controls. Fig. 1 illustrates the problem the operator often faces.

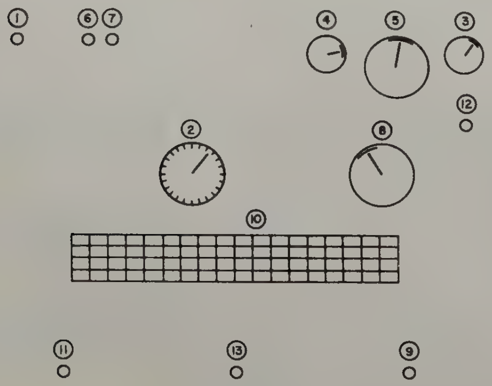


Fig. 1—A panel layout that produces errors of memory.

The numbers on the figure indicate the sequence in which the controls (small circles), meters (large circles), and indicator displays (rectangles) are used. As the numbers indicate, there is no relation between the location of the devices on the panel and the order in which they will be used.

The solution to the problem is quite straightforward. The arrangement of the controls and displays must be altered to reflect the sequence of operations. The standardized sequence should run from left to right and from top to bottom. With such an arrangement the operator will not be forced to rely so heavily on memory or operational manuals.

Errors of Attention

When acting as a monitor, the operator is frequently expected to notice changes in values displayed on a group of meters. Often the meters have null points and the interesting readings are the ones that show deviations from those points. Failure to notice a deviation would be classed as an error of attention.

Fig. 2 illustrates the problem faced by the operator when the null points on his meters have dissimilar orientations. With this arrangement the operator must look at each individual meter to determine if any of them show an out-of-tolerance value. To decrease the possibility of unnoticed error signals, it is desirable to pattern the total display by orienting the null points in the

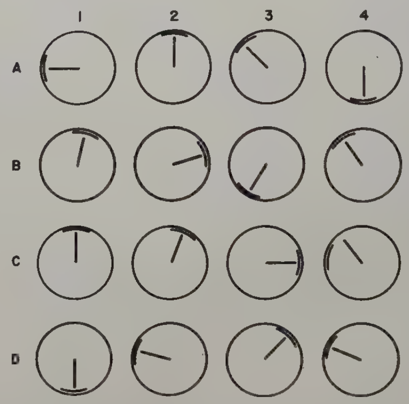


Fig. 2—A difficult checking task.

same direction. The task then becomes that of scanning one or two rows at a time to identify breaks in the patterning. Fig. 3 pictures this principle of display arrangement.

Errors of Identification

An error of identification has been committed when an object is misidentified and then treated as if it were the correct object. There is considerable evidence to suggest that the frequency of errors of identification is much higher than that of any of the other errors. This high frequency may result from the fact that there are so many opportunities for

rrors of identification. Then, too, errors of this ype can be made by the technicians who build the quipment as well as the operators who use it. The roblem of miswired connections is perplexing nany people in electronic industry these days.

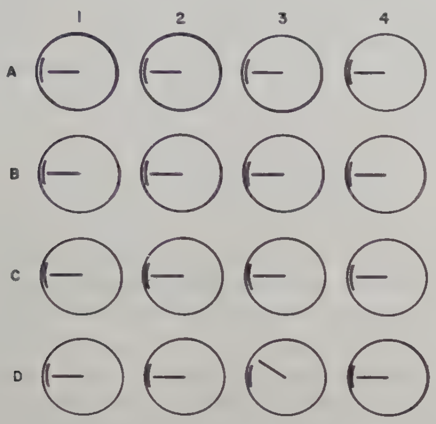


Fig. 3—An improved display for a checking task.

As an illustration of one identification prob- em, Fig. 4 represents a patch board that was ctually used on a piece of checkout equipment. s can be seen, the numbering of the individual

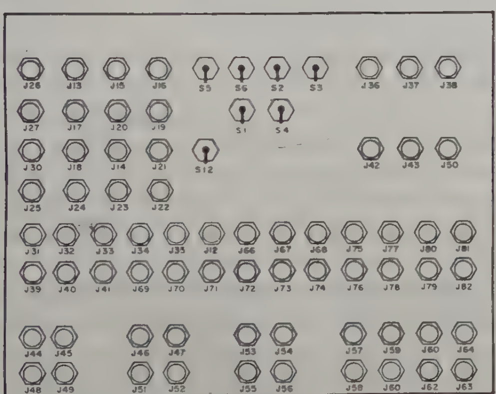


Fig. 4—Identification problems on a patch board.

acks has no regular sequence whatsoever, and nly a partial attempt has been made to break the acks up into individual groups. Fig. 5 shows the atch board that was recommended by human ngineers. On the recommended patch board the equence of the numbers has been preserved and he jacks have been more clearly grouped with the id of black lines. In addition, the subgroups are abeled appropriately. These changes should ertainly reduce errors of identification with his patch board.

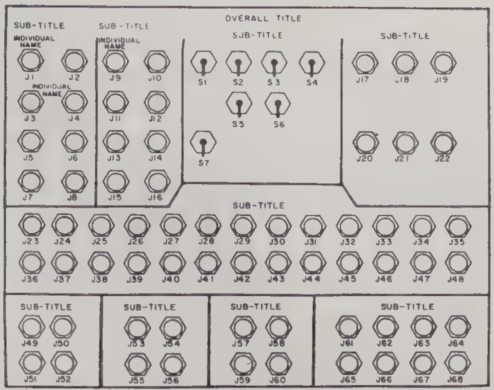


Fig. 5—An improved patch board layout.

Errors of Interpretation

Errors of interpretation occur when the operator misunderstands the meaning of some displayed information and, as a consequence, acts in an in appropriate manner. Fig. 6 pictures a type of dial display that almost assures the occurrence of errors of interpretation. Presumably, one of the two scales is to be used for controlling the amount of "lag," and the other scale is to be used for controlling the amount of "lead." The arrows provide indirect evidence concerning the scale to be used for each purpose. By determining which scale increases in the direction of each arrow, the operator can associate functions with scales. A small empirical test showed, however, that engineers are about as likely to make the incorrect association as the correct one. At the very least, the functional labels should be placed near the scales to which they apply in order to eliminate the ambiguity.

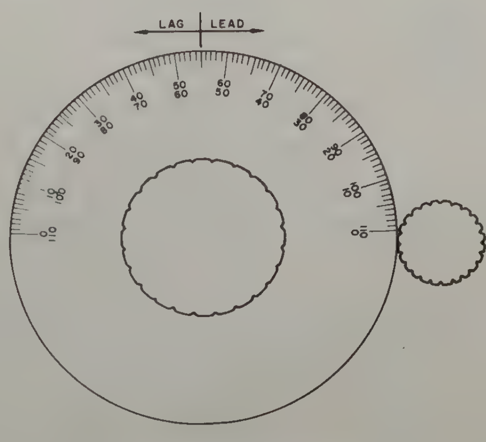


Fig. 6—A dial display associated with errors of interpretation and operation.

Other errors can also be expected with these scales, even after the association problem is solved. On the lower scale, the numbers increase in a counterclockwise direction. It is a safe bet that with such a scale the unnumbered index marks will occasionally be read ten units too high because people have a decided tendency to read scales from left to right. If both scales have to be on the same dial, a warning sign concerning the scale reversal should be installed near the lower scale.

Errors of Operation

For our purposes an error of operation is one in which the control movement is inappropriate to the desired effect. To Fig. 6, I have somewhat arbitrarily added a vernier control know on the right side of the dial in order to illustrate one type of operational error.

Since most vernier knobs operate through a friction drive, a clockwise twist of the vernier knob is sure to produce a scale movement in a direction opposite to that produced with the coarse control knob. Reversal errors in knob operation are the likely result. The replacement of friction drives with appropriate gear drives will be necessary for the achievement of consistent movement relationships.

A Summary Statement

At this point, I want to summarize those principles of human engineering that I have been able only to suggest in this brief survey.

Equipment design should:

1) Facilitate the recall of operational sequences.
2) Exploit the effects of patterning in gaining the operator's attention.
3) Provide redundant cues to object identification.
4) Eliminate the need for display interpretation.
5) Provide for consistent movement relations.

MAN-MAN PROBLEMS

Having examined a few of the human factors problems associated with the use of equipment, we now should consider the problems involved in implementing recommendations that result from our human factors review. At this point our discussion takes on a greater degree of generality, since the problem of gaining acceptance for recommendations is faced by all reliability engineers, regardless of the content of their suggestions.

When the reliability engineer puts down his slide rule and approaches the design engineer with a recommendation, he encounters a new type of problem that he may find more baffling than the technical difficulty he feels he has just licked. He must now deal with a man-man problem in a face-to-face discussion with a design engineer who will probably be polite, but will almost certainly be resistant. The importance of this discussion between the reliability engineer and the design engineer must not be underestimated. The best recommendation ever devised by a reliability engineer will have little effect if it never reaches the hardware stage.

GAINING ACCEPTANCE FOR RECOMMENDATIONS

In discussing methods for improving the acceptance of recommendations, I shall not attempt to suggest specific techniques which the individual reliability engineer might use to influence the individual design engineer. Scientific psychology has really had very little to contribute in the way of improved techniques of this kind. In fact, only in the past few years have psychologists seriously attempted to identify the variables that are related to effective persuasion. Recently, for example, a study was conducted in which husbands and wives were asked to come to an agreement on problems on which they had previously expressed conflicting views [3], the resulting discussions were then analyzed to determine how the conversation of the "winning" husband or wife differed from that of the "loser." One of the conclusions of the study was that the person who talked the most was most likely to come out on top. Perhap this is all that reliability engineers need to do to win their points, but they may have trouble if the design engineers have also read about the same study.

Since the variables affecting persuasiveness have not yet been adequately described, we shall all have to continue to depend largely upon our ow opinions and experience in determining how to de with other people. However, we need not operate completely on our own. We can draw upon the wo of various social scientists to supplement our experience in providing appropriate working conditions that will encourage effective interactions be tween reliability and design engineers.

The conditions that we will want to establish d pend upon the nature of the resistance that we wis to overcome. Examination of the derived needs o both the design engineer and the reliability engineer will provide us with a better understanding c this resistance.

Why Design Engineers Resist Reliability Recommendations

Without attempting a comprehensive discussio of motivation, we can still identify some importar

needs which affect the behavior of design engineers (and the behavior of most other people as well).

There is an abundance of experimental evidence which indicates that a person's needs influence his perception of a situation and his reaction to that situation. Right now we are interested in those needs that help determine the engineer's behavior when faced with a design recommendation from an outside source. Two needs seem to be of particular importance.

Recent surveys show that job security has a high rank in lists of job characteristics most desired by job holders. Just as everyone else, the engineer has a need for job security, and his security is clearly threatened by anyone who recommends a change in a design that the engineer identifies as his own. The reception to such recommendations is likely to be unfavorable.

Furthermore, the engineer is likely to be concerned with his status as an expert, and, again as everyone else, he dislikes to have his work challenged. This need holds with regard to human factors recommendations as well as for recommendations concerning more traditional engineering activity. As one engineer expressed it, "Every engineer also likes to think of himself as a human factors engineer." Having adopted the role of an expert, no one likes to abandon the role without showing some resistance.

In addition to his own needs, the design engineer is influenced by a wide variety of conditions. He may not have a clear idea of the purpose and scope of reliability engineering activities or the services that reliability engineers can perform. He may have received considerable discouragement in the past whenever he himself suggested design changes. Consequently, a change recommended after a design has been frozen is particularly likely to receive brusque treatment. Finally, the engineer is influenced by his supervisor, who may also have negative reactions to reliability recommendations.

Clearly, the design engineer is subject to a number of influences, both internal and external. But what about the reliability engineer? Is he free to operate without regard to these pressures?

Why Reliability Engineers Sometimes Fail

Unfortunately, the reliability engineer is no more able to operate unaffected by personal needs than is the design engineer.

The reliability engineer sees the resistance of the design engineer as a threat to his own job security. In addition, resistance to a recommendation is a serious challenge to the status of the reliability engineer as an expert. As a result, he may develop a defensive attitude in working with design engineers before his opinion is even challenged.

Typically, the reliability engineer also sees himself acting as a critic, a role for which he doesn't particularly care. Furthermore, in this role he expects to alienate most of the people with whom he works. One reliability engineer has indicated his concern by stating: "After a couple years at this job you have so many enemies that you either have to go into another department or to work for a different company."

The reliability engineer may also find himself in an unfavorable position because, although he is a member of a reliability department, he spends most of his time working with people in other departments. In this situation he may find it difficult to identify himself with either department. At the same time the design engineer sees him clearly as an outsider without authority in the design department.

Finally, the reliability engineer has a real need for achievement, a need that is often frustrated by design schedules that make his recommendations obsolete before they are even well formulated.

A Resolution of the Acceptance Problem

Having identified some of the reasons why reliability recommendations may meet resistance, we should try to outline a program that will improve the chances of acceptance. A successful program must reduce resistance and put the reliability engineer in a more favorable position.

For our purposes we will view the lack of acceptance as evidence of a morale problem operating in a two-man group composed of a design engineer and a reliability engineer.

According to one analysis, morale is determined by four factors [1]. To achieve a high level of morale, the members of the group must;

1) Achieve a feeling of group cooperation.
2) Establish a common goal.
3) Have specific tasks that are necessary to the achievement of the goal.
4) See that progress is being made in achieving the goal.

The nice thing about this analysis of morale is that by satisfying the four determinants we can also go a long way toward satisfying the human needs previously described.

Satisfaction of the first determinant of morale requires the establishment of conditions that will foster a genuine feeling of cooperation between the

reliability and design engineers. Somehow we must get across the idea that the reliability engineer is part of the design group in which he works —not an outsider.

In order to satisfy the second determinant, both parties must come to realize the communality of their respective goals. Evidence from studies of social behavior suggests that the similarity in people's attitudes and views is directly related to the frequency of interaction that takes place between them [2]. We should therefore try to provide conditions that will help to satisfy determinants one and two by encouraging frequent interactions between reliability and design engineers. One way in which this may be accomplished is to locate permanently the reliability engineer in the same work space with the engineers with whom he will work, where opportunities for interaction will be at a maximum.

To achieve the greatest benefit from this arrangement, it must be clear that the reliability engineer has a status equal to that of the people with whom he most frequently deals. Channels of authority must also be established. Practically, this means, among other things, that the reliability engineer must have as desirable a work space as his design counterpart and that the design supervisor must clearly establish the authority of the reliability engineer in the departmental organization.

It should be noted that these procedures will also aid the reliability engineer in identifying with the group with which he works. The status of the reliability engineer as an expert will be improved as well. Once the reliability engineer is accepted as part of the design team, the unique contributions of both the reliability and design engineers will be seen to be directed toward the same goal —the design of the best possible piece of equipment.

If the limited size of the reliability staff makes it necessary for reliability engineers to work with several different groups of design engineers, it may not be desirable to locate individual reliability engineers within any one particular group. Under these circumstances, cooperation may be more difficult to achieve. There are, however, additional techniques, which should be employed in any case. These techniques, which are described in the following discussion, will help to compensate for the loss.

Thus, final achievement of cooperation and goal alignment will depend on the development of conditions that will enable the reliability engineer to accept responsibility as a participant in initial design stages rather than as a critic whose inputs are often too late. To achieve the desired state, the reliability engineer must shift the major share of his attention to projects that are in their initial stages even if he must, as a result, neglect some of the fire drills that are in continual rehearsal. The benefits from such a shift in emphasis have a large potential. A systematic program for reliability efforts is almost certain to produce better results than a haphazard program. Furthermore, the acceptance of reliability recommendations will be enhanced by the resulting satisfaction of the human needs previously discussed. As the reliability engineer becomes a responsible contributor during initial design stages, his contributions become less of a threat to the design engineer. The reliabilty engineer also benefits. He no longer has to resent his role as a critic who, because of the nature of his role, must expect to alienate the people with whom he works. Finally, as he sees the influence of his recommendations appearing in finished equipment, the reliability engineer's need for achievement should be gratified.

During its early stages of development the cooperative spirit can be strengthened by a mutual program of cross education. From the reliability end, the educational program should consist of printed design bulletins containing general recommendations for the improvement of system reliability. These bulletins should be supplemented with frequent seminars during which the scope of the reliability program is indicated and methods of cooperation are discussed.

The methodology of cooperation must not be slighted, for the satisfaction of morale determinant number three requires the establishment of specific tasks related to the accomplishment of the mutual goal. It will not be sufficient to establish a cooperative atmosphere without also establishing detailed procedures for interaction and exchange of ideas. The agreed upon techniques will undoubtedly differ from one situation to another, but they should be products of a group decision on the part of the reliability and design engineers.

The final determinant of morale requires that cooperating engineers must see that they are making progress toward their joint goal of increased reliability. To be effective, this feedback should be as immediate and specific as possible. What is needed are numbers that reflect reductions in the frequency with which failures are experienced as a result of equipment modifications initiated to improve reliability. As the collection of failure data itself becomes reliable, special attempts should be made to use the data to supply the desired knowledge of results. Until

this step is taken, the requirements of our four determinants will not be fulfilled.

CONCLUSIONS

Although the results achieved will depend on the particular people involved, it is clear that attention to human factors can improve the reliability attained in the design and operation of complex equipment.

Human factors are important in the operation of equipment for the obvious reason that humans are doing the operating. When equipment design is inadequate, human operators are prone to errors that may seriously reduce the inherent reliability of the components of the equipment. Design modifications can curtail the opportunities for these man-machine errors.

Human factors are also important to equipment design because the acceptance of reliability recommendations is affected by the personal needs that influence the behavior of both design and reliability engineers. The lack of effective cooperation between reliability and design engineers may be viewed as a symptom of low morale. Resolution of the morale problem requires the establishment of working conditions that will help the reliability and design engineers to:

1) Achieve a feeling of group cooperation.
2) Establish a common goal.
3) Decide on specific tasks that will lead to the goal.
4) See that progress is being made toward the goal.

The satisfaction of these four determinants of morale can be at least partially achieved through appropriate organizational arrangements and a shift in the attention of the reliability engineer to the earliest design stages.

REFERENCES

[1] M. L. Blum, "Industrial Psychology and Its Social Foundations" Harper and Brothers, New York, N. Y.; 1949.
[2] G. C. Homans, "The Human Group," Harcourt, Brace and Co., Inc., New York, N. Y.; 1950.
[3] F. L. Strodtbeck, "Husband-wife interaction over revealed differences," Amer. Social. Rev., vol. 16; 1951.

ESTIMATION FROM LIFE TEST DATA*

BENJAMIN EPSTEIN†

We first discuss problems of estimation where the life X is described by the pdf

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0, \quad \theta > 0. \qquad (1)$$

Case I: Life testing is discontinued after a fixed number r of items have failed. Items on test may or may not be replaced. The number of items initially on test is n.

The "best" estimate of the mean life θ is given by

$$\hat{\theta}_{r,n} = T_{r,n}/r, \qquad (2)$$

where $T_{r,n}$ is the accumulated life on test until the rth failure occurs. The observed failure times are $x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{r,n} \leq \cdots$. For simplicity of notation, we will suppress the subscript n.

If testing is terminated after $r (\leq n)$ failures have occurred, then in the nonreplacement case:

$$T_r = \sum_{i=1}^{r} x_i + (n - r)x_r. \qquad (3)$$

In the replacement case (where r is now unrestricted),

$$T_r = n\, x_r. \qquad (4)$$

The probability density function of $\hat{\theta}_r$, in either the replacement or nonreplacement case, is given by

$$f_r(y) = \frac{1}{(r-1)!} \left(\frac{r}{\theta}\right)^r y^{r-1} e^{-ry/\theta}, \quad y > 0 \qquad (5)$$

$$= 0, \quad \text{elsewhere.}$$

From (5) it follows that

$$\frac{2r\,\hat{\theta}_r}{\theta} = \frac{2\,T_r}{\theta} \text{ is distributed as } \chi^2(2r) \qquad (6)$$

(*i.e.*, as chi square with 2r° of freedom).

From (6) it follows that a two-sided $100(1-\alpha)$ per cent confidence interval for θ is given by

$$\frac{2\,T_r}{\chi^2_{\frac{\alpha}{2}}(2r)} < \theta < \frac{2\,T_r}{\chi^2_{1-\frac{\alpha}{2}}(2r)}, \qquad (7)$$

where $\chi^2_{\frac{\alpha}{2}}(2r)$ is the upper $\frac{\alpha}{2}$ percentage point of $\chi^2(2r)$ and $\chi^2(2r)$ and $\chi^2_{1-\frac{\alpha}{2}}(2r)$ is the lower $\frac{\alpha}{2}$ percentage point of $\chi^2(2r)$. Similarly a one-sided $100(1-\alpha)$ per cent confidence interval for θ is given by

$$\theta > \frac{2\,T_r}{\chi^2_{\alpha}(2r)}, \qquad (8)$$

where $\chi^2_{\alpha}(2r)$ is the upper α percentage point of $\chi^2(2r)$.

Put into words, $100(1-\alpha)$ per cent of assertions of the kind made in (7) and (8) will be correct.

Example 1

20 electron tubes are placed on test. A tube which fails is replaced at once by a new tube. The fifth failure is observed to occur 407 hours after the start of the life test.

Estimate the mean life θ and give one and two-sided 95 per cent confidence intervals for θ.

Solution: We are dealing with a replacement situation with $n = 20$, $r = 5$, $x_5 = 407$. The total life observed is, according to (4), given by $T_5 = 20x_5 = 20(407) = 8140$. Thus it follows from (2) that $\hat{\theta} = T_5/5 = 8140/5 = 1628$ hours. To find a two-sided 95 per cent confidence interval for θ, we use (7) with $\chi^2_{0.025}(10) = 20.483$ and $\chi^2_{0.975}(10) = 3.247$. Substituting in (7) we get the two-sided 95 per cent confidence interval $795 < \theta < 5014$. To find a one-sided 95 per cent confidence interval we use (8) with $\chi^2_{0.05}(10) = 18.307$. Substituting in (8) we get the one-sided 95 per cent confidence interval $\theta > 889$ hours.

Frequently we are not only interested in estimating θ, but also in estimating a quantity x_p where x_p is that life such that

$$\Pr(X > x_p) = p. \qquad (9)$$

*This paper was presented at the First Annual Bay Area Reliability Seminar, Menlo Park, Calif., February 19, 1959. This work was done with the partial support of the Office of Naval Research. This material is covered in detail in Technical Report No. 4, ONR Contract Nonr-2163(...).

†Wayne State University, Detroit, Mich.

For the exponential distribution

$$\hat{x}_p = \hat{\theta} \log \frac{1}{p} \,.^{1} \tag{10}$$

Maximum likelihood estimates and $100(1 - \alpha)$ per cent confidence intervals for x_p are given by

$$x_p = \theta_r \log \frac{1}{p} \,, \tag{11}$$

$$\frac{2\,T_r \log \frac{1}{p}}{\chi^2_{\frac{\alpha}{2}}(2r)} < x_p < \frac{2T_r \log \frac{1}{p}}{\chi^2_{1-\frac{\alpha}{2}}(2r)} \tag{12}$$

in the two-sided case and

$$x_p > \frac{2\,T_r \log \frac{1}{p}}{\chi^2_{\alpha}(2r)} \tag{13}$$

in the one-sided case.

Eq. (13) can be interpreted as follows.

We can be $100(1 - \alpha)$ per cent confident of the truth of the assertion that the probability of surviving $\tau = 2\,T_r \log \frac{1}{p}/\chi^2_{\alpha}(2r)$ time units is $\geq p$.

This is a tolerance interval statement in that we can be $100(1 - \alpha)$ per cent confident of the correctness of the assertion that the fraction of items surviving τ or more time units is $\geq p$. Putting the last statement into reliability language, we can be $100(1 - \alpha)$ per cent confident that the reliability over $[0,\tau]$ is $\geq p$.

Example 2

Given the data in Example 1, estimate $x_{.9}$, where $x_{.9}$ is such that

$$\Pr(X > x_{0.9}) = 0.9 \tag{14}$$

(*i.e.*, the probability of surviving for $x_{0.9}$ hours is 0.9). Give one- and two-sided 95 per cent confidence intervals for $x_{0.9}$.

Solution: $\log \frac{1}{p} = \log \frac{1}{.9} = .1054$. Hence, substituting in (11) we get

$$x_{0.9} = (1628)(0.1054) \tag{15}$$
$$= 172 \text{ hours.}$$

Substituting in (12) we get the two-sided 95 per cent confidence interval: $83.8 < x_{0.9} < 528$ and substituting in (13) we get the one-sided 95 per cent confidence interval $x_{0.9} > 93.7$ hours.

[1]All logarithms used in this paper are natural logarithms.

Footnote 1

Example 3

Given the data in Example 1, find a number τ such that we can assert with 95 per cent confidence that at least 90 per cent of the items in the population survive τ hours. (Note that this is a tolerance statement. It can also be given a reliability interpretation.)

Solution: We have noted above that one-sided $100(1 - \alpha)$ per cent confidence statements regarding x_p are also tolerance statements in which we can have $100(1 - \alpha)$ per cent confidence. Hence it follows from the solution to Example 2 that we can assert with 95 per cent confidence that at least 90 per cent of the items in the population survive $\tau = 93.7$ hours.

We are frequently interested in making point and interval estimates about the probability that the item survives a preassigned length of time t^*. Denoting this by p_{t^*}, we have

$$p_{t^*} = \Pr(X > t^*) = e^{-t^*/\theta} \,. \tag{16}$$

It is obvious how one can make point and interval estimates of p_{t^*} from the corresponding formulae for $\hat{\theta}_r$ [see (2), (7), and (8)]. In particular, a one-sided $100(1 - \alpha)$ per cent confidence interval for p_{t^*} is given by

$$p_{t^*} > e^{-\chi^2_{\alpha}(2r)t^*/2\,T_r} \,. \tag{17}$$

The question may be asked: how large should the observed T_r be in order that we be $100(1 - \alpha)$ per cent confident that

$$p_{t^*} = e^{-t^*/} \geq \gamma \;? \tag{18}$$

From (17) this implies that

$$\exp[-\chi^2_{\alpha}(2r)t^*/2\,T_r] \geq \gamma \tag{19}$$

or

$$T_r \geq \chi^2_{\alpha}(2r)t^*/2 \log \frac{1}{\gamma} \,. \tag{20}$$

The interpretation of (20) is as follows.

If the total life observed in getting r failures exceeds $\chi^2_{\alpha}(2r)t^*/2 \log \frac{1}{\gamma}$, then we can be $100(1-\alpha)$ per cent confident of the assertion that the probability of surviving time t^* is $\geq \gamma$. In reliability considerations we can replace the words "reliability over a time interval of length t^* is $\geq \gamma$."

Example 4

Given the data in Example 1, make one- and two-sided 95 per cent confidence statements for the probability of surviving $t^* = 100$ hours.

Solution: The maximum likelihood estimate of p_{t*}, the probability of surviving $t^* = 100$ hours, is given by

$$\hat{p}_{t*} = e^{-100/1628} = e^{-.0614} = 0.9404.$$

Similarly, a two-sided 95 per cent confidence interval for p_{t*} is given by

$$(e^{-100/795}, e^{-100/5014}) = (0.8817 < p_{t*} < 0.9802).$$

A one-sided 95 per cent confidence interval for p_{t*} is given by substituting in (17). This gives us

$$p_{t*} > 0.8936.$$

We can be 95 per cent confident of the assertion that the probability of surviving 100 hours (reliability over the time (0,100)) is ≥ 0.8936.

Example 5

The total life observed in obtaining 5 failures is 9205 hours. On the basis of this information, can we be 95 per cent confident that the probability of surviving (reliability) for a time $t^* = 100$ is ≥ 0.90?

Solution: From (20) it is known that in order to be 95 per cent confident that the probability of surviving for a time t^* is ≥ 0.9, it is necessary that the total observed life

$$T_5 \geq \chi^2_{0.05}(10)100/2 \log \frac{1}{0.9} = 8689.$$

Since the total life observed in obtaining 5 failures is 9205, we can answer in the affirmative, *i.e.*, we can be 95 per cent confident that the probability of surviving for a time $t^* = 100$ is ≥ 0.90.

Case II: Underlying distribution is exponential. The life test is discontinued after a fixed amount of total life T has elapsed. Items under test may[2] or may not be replaced.

In what follows, let r = number of items which fail in $[0, T]$; then some formulas of interest are:

Two-sided $100(1 - \alpha)$

per cent confidence interval for θ

$$\frac{2T}{\chi^2_{\frac{\alpha}{2}}(2r + 2)} < \theta < \frac{2T}{\chi^2_{1-\frac{\alpha}{2}}(2r)} . \tag{21}$$

One-sided $100(1 - \alpha)$

per cent confidence interval for θ

$$\theta > 2\,T/\chi^2_{\alpha}(2r + 2) . \tag{22}$$

One-sided $100(1 - \alpha)$

per cent confidence interval for the quantity

$$x_p = \theta \log \frac{1}{p} .$$

$$x_p > 2\,T \log \frac{1}{p} / \chi^2_{\alpha}(2r + 2) . \tag{23}$$

If we define τ as

$$\tau = 2\,T \log \frac{1}{p} / \chi^2_{\alpha}(2r + 2) , \tag{24}$$

then we can assert with $100(1 - \alpha)$ per cent confidence that at least 100p per cent of the items survive for a length of time τ. Putting the last statement into reliability language we can be $100(1 - \alpha)$ per cent confident of the truth of the assertion that the reliability over $[0, \tau]$ is $\geq p$.

Example 6

30 items are placed on test. Items which fail are replaced. The life test is stopped after 100 hours have elapsed. Five failures were observed in the course of the experiment. Assuming that the underlying distribution of life is exponential, find one- and two-sided 95 per cent confidence intervals for θ.

Solution: In this problem the fixed amount of total life observed is $T = nt^* = 30(100) = 3000$. Substituting in (21) and using $\chi^2_{0.025}(12) = 23.337$ and $\chi^2_{0.975}(10) = 3.247$, one gets the two-sided 95 per cent confidence interval,

$$257 < \theta < 1848 .$$

Substituting in (22) and using $\chi^2_{0.05}(12) = 21.026$, we get the one-sided 95 per cent confidence interval

$$\theta > 285.$$

Example 7

Given the data in Example 6, estimate τ so that we will be 95 per cent confident that the probability of surviving τ hours is at least 0.9. Substituting in (24), and using $T = 3000$, $r = 5$, $\alpha = 0.05$, $p = 0.9$, we get

$$\tau = \frac{6000}{21.026}(0.1054) = 30.1 .$$

On the basis of the data we can be 95 per cent

[2] In the important special case where n items are tested, with replacement, for a length of time t^*, $T = nt^*$.

confident that the probability of surviving τ = 30.1 hours is ≥ 0.9.

Case III: n items are placed on life test for a time t^*. At the end of this time one counts the number of items that have failed in $[0,t^*]$. *Items that fail are not replaced.*

In what follows let r = number of observed failures. Then we can make the following *nonparametric statement.*

We can assert with $100(1 - \alpha)$ per cent confidence that at least 100b per cent of the population survives for a length of time t^* with b given by

$$b = \frac{1}{1 + \left(\frac{r+1}{n-r}\right) F_\alpha\,(2r + 2,\ 2n - 2r)} . \qquad (25)$$

Put in reliability language, we are $100(1 - \alpha)$ per cent confident of the assertion that the reliability over $[0,t^*]$ is $\geq$ b.

In the *particular* case where the underlying distribution is *exponential*, a one-sided $100(1 - \alpha)$ per cent confidence interval for θ is given by

$$\theta > \frac{t^*}{\log\left\{1 + \left(\frac{r+1}{n-r}\right) F_\alpha(2r + 2, 2n - 2r)\right\}} . \qquad (26)$$

In (25) and (26), $F_\alpha(2r + 2,\ 2n - 2r)$ is the upper α percentage point of the $F(2r + 2,\ 2n - 2r)$ distribution.

Example 8

20 items are placed on life test for 100 hours. Two items fail before this time. Items which fail are not replaced.

a) Make a nonparametric one-sided 95 per cent confidence statement about the probability of surviving 100 hours.

b) If the underlying distribution is exponential, find a one-sided 95 per cent confidence interval for the mean life θ.

Solution: a) In this problem n = 20, r = 2, α = 0.05, t^* = 100. Since $F_{0.05}(6,36) = 2.36$, it follows from (25) that b = 0.718. Hence we can make the following nonparametric statement.

We are 95 per cent confident of the assertion that the probability of surviving 100 hours (reliability over 100 hours) is $\geq$ 0.718.

b) Substituting in (26) we get the one-sided 95 per cent confidence interval, $\theta > 302$.

Example 9

Ten thousand one-hour missions are carried out. Ten failures are observed. Make a one-sided 95 per cent confidence statement about the reliability (probability of success) in a one-hour mission.

Solution: In this problem n = 10,000, r = 10. Substituting in (25) we get b = 0.9983. We can be 95 per cent confident that the reliability in a one-hour mission is $\geq$ 0.9983.

Remark: In carrying out the computations we use the fact that since n is large $F_{0.05}(22;19980) \sim F_{0.05}(22,\infty) = 1.54$.

Example 10

Suppose that 10,000 one-hour missions are carried out and that no failures are observed. Find a one-sided 95 per cent confidence interval for the probability of mission survival.

Solution: Substitute in (25) with n = 10,000, r = 0. $F_{0.05}(2;20{,}000) \sim F_{0.05}(2,\infty) = 3.00$. Hence, b = 0.9997. We can have 95 per cent confidence in the assertion that the true probability of mission survival is $\geq$ 0.9997.

A CUSTOMER LOOKS AT THE RELIABILITY PROGRAM ACTIVITIES*

H. R. POWELL†

INTRODUCTION

The Space Technology Laboratories of Los Angeles has the responsibility for systems engineering and technical direction of all contractors on the Air Force Ballistic Missile Programs. As such, we work in close association with the members of the Ballistic Missile Division of the United States Air Force, in establishing, implementing, and directing the various activities of the contractors who are working on these programs. However, I have not always represented the military-customer side of the picture. I was for many years a contractor before becoming associated with the present programs. Therefore, my comments will stem from the combined background of both a contractor and a military customer. I will attempt, therefore, to put the requirements of the vast and complex military and Department of Defense organizations into the language of a contractor. The remarks contained in the text of this paper will be a combination of what I know about the requirements and policies of the Department of Defense, the United States Air Force, and the policies of my own company, STL.

To say that the Department of Defense and the Military Services have a high interest in reliability would indeed be an understatement. It is generally recognized by these agencies that the reliability of a system has tremendous implications in terms of overall weapon-system effectiveness, in logistics and manpower, and in terms of forced sizing of the various component arms of our total military organizations. To say that these people are concerned by the overall reliability record established in the past is also putting it mildly. I would like to quote from the report of the Department of Defense Ad Hoc Committee for Guided Missile Reliability, dated April, 1958: "A significant number of guided missile weapon systems have failed to achieve the required reliability in the time scheduled and at the estimated cost. Consequently, military planning is upset, funds are expended that might have been usefully applied to more productive projects, and the operational use of missile systems is postponed, thereby diminishing our nation's military strength."[1]

The realization of the need for reliability suddenly burst upon us when this country embarked on large-scale missile programs to be used in our offensive and defensive arsenals, and the emphasis on reliability has increased over the past ten years. It is felt, however, that reliability technology and achievements have not generally kept pace with the requirements. Therefore, there has been an evolutionary trend toward the concept of spelling out reliability requirements in more specific terms in contracts. Contractors can expect to see increasingly stringent reliability requirements spelled out in contracts in numerical forms. This concept is now part and parcel of the BMD/STL Reliability policy for the ballistic missile contractors. Similar requirements are beginning to appear in the contracts of the other Services.

This procedure has several advantages: It tells the management of the company just what they are expected to deliver in the way of systems reliability; it gives the contractors' own reliability organizations something definite to hang their hat on and to use as a basis for carrying out their activities; and it gives the military, the customer, better assurance that the systems he buys will be usable.

A SYSTEMS-RELIABILITY PROGRAM

In order to spell out more clearly what a contractor is expected to do about reliability, it has become customary to identify the kinds of activities which are considered valid expenditures of manpower and money, a measure essential to the ultimate achievement of the required reliability. The Ballistic Missile Division and the Space Technology Laboratories consider the following

*This paper was presented at the First Annual Bay Area. Reliability Seminar, Menlo Park, Calif., February 19, 1959.
†Space Technology Labs., Rano-Wooldridge Corp., Los Angeles, Calif.

[1] Dept. of Defense Ad Hoc Comm. for Guided Missile Reliability report, publ. by Office of William Holaday, Director of Guided Missiles, Dept. of Defense, Washington, D. C.

ontractor activities, each described briefly, to e essential to the carrying out of a balanced and omplete reliability program: (Of course, the ist could be lengthened or shortened through a ubdivision of some of these program elements, r a combination of some, but these would be es- entially variations of this same basic reliability rogram.)

) Environmental Conditions Determination

Missile environment measurements should be eviewed, and interactive effects of the contrac- or's hardware on missile environment should be stimated. Environmental conditions for assem- ly and parts specifications should be determined. nformation on environmental conditions should e disseminated to designers, specification riters, and test planners.

) Reliability Apportionment

Detailed reliability objectives should be estab- ished by numerically apportioning the contrac- or's system reliability objective among the com- onent assemblies and parts. The objectives hould be disseminated for use in design guid- nce, in evaluations and comparisons, in planning f test programs, and in pinpointing problem reas.

) Reliability Indoctrination

The reliability program requirements and rocedures should be explained and "sold" to ompany personnel and to vendors. Measures uch as a training program for engineers should e instituted.

) Parts Approval Verification

The component parts engineering effort (wheth- r centralized or not) should be monitored. Qual- fication tests and performance data on parts hould be reviewed, coordinated, and dissemi- ated. Information sources such as preferred arts lists, data files, and vendor ratings should e maintained.

) Specifications Review

Specification writers should be assisted with egard to reliability objectives and statistical olerance considerations. Specifications should e reviewed to assure that reliability will not be nduly compromised in such matters as reliabil- ty-performance trade-offs.

) Design Review

Designers should be assisted in such matters s environment problems, tolerances, component application, developmental marginal checking, consideration of effects of production tooling methods on reliability, and human use factors. Designs should be reviewed for such reliability factors as adequate safety margins, provision for preventive maintenance, and appropriate redun- dancy.

7) Failure Reporting Surveillance

Procedures should be established and main- tained for reporting individual failures in plant and field operations involving prototype and pro- duction hardware. Individual failure reports should receive engineering analysis and be dis- tributed to design or production activities for prompt correction of troubles. Follow-up should be instituted to assure that failures are corrected. A card file should be maintained on failure-report data. Summary reports should be prepared from the card file and distributed, so that problem areas and progress may be defined.

8) Statistical Test Planning

Test engineers should be assisted so that opti- mum consideration is given to environmental test conditions, reliability objectives, and statistical design of experiments (particularly with regard to sample size, stress level, and arrangement of tests). Test plans should be reviewed to assure that testing will be sufficiently comprehensive to allow detection of important modes of failure and to provide a basis for effective evaluation. "Op- erating characteristic" curves should be pre- pared to define the effectiveness of the planned tests.

9) Statistical Test Evaluation

All test reports should be analyzed and evalu- ated for information on hardware capabilities and weaknesses. Data and results should be explained to designers, and also stored and cross-filed for reference so that accumulated information on particular designs may be readily located at any time. Information should be included on succes- ses as well as failures.

10) Quality Control Coordination

The quality control effort should be coordinated with regard to such matters as reliability objec- tives, production process control, and inspection test procedures designed to show up incipient failures.

11) Program Data Evaluation

All information and data obtained in the pro- gram should be organized, analyzed, and evaluated

with regard to reliability. Estimates of achieved reliability should be made and projected to operational use. Reports should be issued to define problem areas and report progress. A list of critical assemblies and parts should be maintained. A special effort should be made to "close the loop" by feeding information back into the organization promptly and at points where it is needed.

12) Vendor Control

The contractor shall take steps to ascertain through proper tests and surveillance that parts and devices supplied by vendors and subcontractors are adequate for their intended application in the contractors' equipment. These measures shall include tests to demonstrate design capability and to provide a continuous monitoring of the vendor's quality control and product improvement programs.

13) Flight Test Planning

Plans and specifications for missile flight testing should be reviewed from the point of view of obtaining as much information pertinent to reliability as possible without causing undue compromise or interference with other flight test objectives during their R & D program. Instrumentation should be carefully reviewed as to adequacy of design and ability to yield data of required quality. Special attention should be given to the question of whether or not certain telemetering channels should be commutated when the reliability of a missile subsystem is being evaluated by the operating time-to-failure criteria. For flights during the latter stages of the R & D program and post-IOC flights, inclusion of reliability objectives and flight test planning is even more important, in the sense that flight-test data on missiles of operational or near-operational design are especially significant in making estimates of future reliability of operational systems.

14) Analysis of Test and Flight Failures

All failures resulting from environmental, factory, field, and flight test should be analyzed by people in the reliability organization to determine the significance of the failures in a reliability sense. The analysis should be on both a statistical and engineering basis. An attempt should be made to determine causes of the failures and, where feasible, whether or not the deficiency was due to design, quality control, or human factor. Where appropriate, these analyses should be fed back through the failure-reporting system described in section 7, above.

15) Determination of Corrective Action

Analysis of test results and failures should have two primary purposes: To determine the need for corrective action, and to establish at least a recommendation as to the nature of the corrective action required.

16) Corrective Action Follow-up

This constitutes one of the most critical of all reliability activities. Procedures should be set up which will allow for a routine follow-up of corrective actions recommended or being acted upon, and should contain a system of checks and balances to assure that the required corrective actions have not been forgotten or ignored by those responsible for implementing the final action.

It is recognized that the interpretation and implementation of these requirements may vary somewhat between contractors, depending upon the status of existing contracts, level and severity of the reliability requirement on that contractor, state of the art of the technical field in which he is working, and the contractor's own organization, policies, and procedures. However, he is expected to show that these activities are being carried out in one form or another.

The question of contractor organization is considered a very important point. Several things are expected of the contractor in this respect. First of all, he must have a separate and distinct reliability organization adequately staffed with technically competent people. Secondly, the organization should be vested with certain in-line responsibilities for carrying out a meaningful program. Finally, it should report sufficiently high up in the company's organization to give it authority and management backing.

GOVERNMENT REPORTS ON RELIABILITY

Two of the most significant of the recent documents to come out of the Department of Defense on the subject of reliability have been the reports of the Advisory Groups on Reliability of Electronic Equipment, called AGREE, and the report of the Ad Hoc Committee on Guided Missile Reliability. It should be noted that many of the activities described above coincide with activities recommended by AGREE and the ACGMR. The AGREE and ACGMR reports were intended to complement each other and not to duplicate each other's work. The AGREE report is a collection of detailed technical procedures for dealing with many of the specific problems which arise in carrying out an active reliability program. The ACGMR report,

on the other hand, is considered to be more of a management-type document to be used by both contractor-management and military-management in guiding the overall efforts in the reliability programs under their cognizance. Thus, the AGREE and ACGMR reports form an excellent team; one the detailed technical procedures to be used in reliability; the other the management procedures to be used. In general, it has been found that those contractors who conscientiously carry out programs of the kind described above show very promising results.

Well known, by now, is the investigation which was carried out during the latter half of 1958 by a scientific study group under the auspices of the House Appropriations Subcommittee. (Mahon Committee.) This investigation apparently stemmed from a feeling of apprehension on the part of our national leaders as to whether or not the vast sums of money which are being spent on our missile and other weapons systems programs are producing weapon systems which are reliable. The results of the Study Groups' activities have not been made public at this time. However, we can anticipate that the results of this investigation and study will be seen in increased pressure on everyone from the Department of Defense down through the Military Services and eventually to the contractors, for the attainment of a higher reliability than has been demonstrated in the past.

CONCLUSION

1) The push is on for higher reliability and increased efforts on the part of contractors.
2) The military customers do not want lip service alone.
3) Contractors are expected to have a distinct and recognizable reliability organization.
4) The organization to be effective must have in-line responsibility and authority.
5) There must be an actual reliability program in being with definite reliability activities being carried out, not just a paper program which rationalizes the problem.
6) There must be management support for reliability.
7) The military customers encourage original work on the part of the contractors to develop their techniques for analyzing and understanding the reliability problem and for generating solutions to the problem.
8) The military customers encourage more active interchange of information between contractors, not just through symposia and seminars, although these are very valuable, but also through more direct contact and discussion with each other.

CORRECTION TO "MODULE PREDICTION"

George Hauser, author of "Module Prediction," which appeared on pages 53-63 of the June, 1959 issue of these TRANSACTIONS, has requested that the following corrections be made to his paper.

Eq. (1), page 56, should read

$$\frac{\partial A}{A} = \frac{1}{1+\mu\beta} \; \frac{\partial \mu}{\mu}$$

Eq. (2), page 56, should read

$$\frac{\partial A}{A} = \frac{-\mu}{1+\mu\beta} \; \frac{\partial \beta}{\beta}$$